ISSN 1645-6726

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

REVSTAT
Statistical Journal

Volume 15, No.1
January 2017

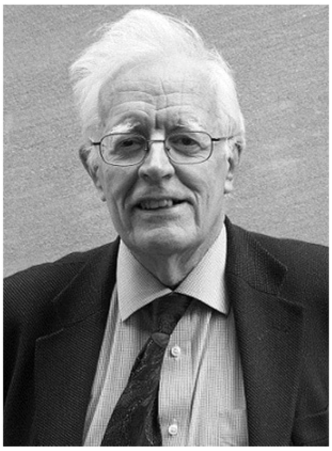
REVSTAT
Statistical Journal

Catalogação Recomendada

REVSTAT. Lisboa, 2003-
Revstat : statistical journal / ed. Instituto Nacional de Estatística. - Vol. 1, 2003- . - Lisboa I.N.E., 2003- . - 30 cm
Semestral. - Continuação de : Revista de Estatística = ISSN 0873-4275. - edição exclusivamente em inglês
ISSN 1645-6726

CREDITS

- **PUBLISHER**
  - *Instituto Nacional de Estatística, I.P. (INE, I.P.)*
    Av. António José de Almeida, 2
    1000-043 LISBOA
    PORTUGAL
    Tel.: + 351 21 842 61 00
    Fax: + 351 21 845 40 84
    Web site: http://www.ine.pt
    Customer Support Service
    (National network) : 808 201 808
    Other networks: + 351 218 440 695
- **COVER DESIGN**
  - *Mário Bouçadas, designed on the stain glass window at INE by the painter Abel Manta*
- **LAYOUT AND GRAPHIC DESIGN**
  - *Carlos Perpétuo*
- **PRINTING**
  - *Instituto Nacional de Estatística, I.P.*
- **EDITION**
  - *150 copies*
- **LEGAL DEPOSIT REGISTRATION**
  - *N.º 191915/03*
- **PRICE [VAT included]**
  - *€ 9,00*



Congratulations, Sir David Cox!

International Prize of Statistics

Sir David Cox has been named the first recipient of the International Prize of Statistics created by the American Statistical Association, the Institute of Mathematical Statistics, the International Biometric Society, the International Statistical Institute and the Royal Statistical Society.

The International Prize in Statistics Foundation announcement states that this award specifically recognizes Sir David Cox for his 1972 paper in which he developed the proportional hazards model that today bears his name. The announcement (http://www.statprize.org/pdfs/Press-Release-International-Prize Winner.pdf) recognizes as well that Sir David Cox "is a giant in the field of Statistics".

REVSTAT congratulates Sir David Cox for this well deserved attribution of the prize, and takes the opportunity to thank him for invaluable guidance as member of the editorial board since the beginning, and for his many contributions to REVSTAT publication standards.

INDEX

Abstracted/indexed in: *Current Index to Statistics, DOAJ, Google Scholar, Journal Citation Reports/Science Edition, Mathematical Reviews, Science Citation Index Expanded®, SCOPUS* and *Zentralblatt für Mathematic.*

REVSTAT – Statistical Journal
Volume 15, Number 1, January 2017, 1–23

OPTIMIZING THE SIMPLE STEP STRESS ACCELERATED LIFE TEST WITH TYPE I CENSORED FRÉCHET DATA

Authors: Nooshin Hakamipour
– Department of Statistics, Amirkabir University of Technology, (Tehran Polytechnic), Iran
nooshin.hakami@aut.ac.ir

Sadegh Rezaei
– Department of Statistics, Amirkabir University of Technology, (Tehran Polytechnic), Iran
srezaei@aut.ac.ir

Received: April 2014 Revised: February 2015 Accepted: May 2015

Abstract:

- In this paper, we propose an optimization for the simple step stress accelerated life test for the Fréchet distribution under type I censoring. The extreme value distribution has recently become increasingly important in engineering statistics as a suitable model to represent phenomena with extreme observations. One probability distribution, that is used to model the maximum extreme events, is the Fréchet (extreme value type II) distribution. A log-linear relationship between the Fréchet scale parameter and the stress are assumed. Furthermore, we model the effects of changing stress as a cumulative exposure function. The maximum likelihood estimators of the model parameters are derived. By minimizing the asymptotic variance of the desired life estimate and the reliability estimate, we obtain the optimal simple step stress accelerated life test. Finally, the simulation results are discussed to illustrate the effect of the initial estimates on the optimal values.

Key-Words:

- *Fréchet distribution; log-linear relationship; maximum likelihood estimator; optimal design; reliability; step stress accelerated life test; type I censored data.*

AMS Subject Classification:

- 90B25, 62P30.

1. INTRODUCTION

Nowadays, manufacturers face strong pressure to rapidly develop new, higher technology products, while improving the productivity. This has motivated the development of methods such as concurrent engineering and encouraged wider use of the designed experiments for product and process improvement. The requirements for higher reliability have increased the need for more up front testing of the materials, components, and systems. This is in line with the modern quality philosophy for producing the high reliability products: achieve high reliability by improving the design and manufacturing processes; move away from reliance on inspection (or screening) to achieve high reliability [12].

Estimating the failure time distribution of components including high reliability products is particularly difficult. Most modern products are designed to operate without failure for years, decades, or longer. Thus, a few units will fail or degrade in a test under the normal conditions. For example, the design and construction of a communication satellite, may allow only 8 months to test the components that are expected to be in service for 10 or 15 years. A method for obtaining information on the life distribution of a product in a timely fashion, is to test it on an unusually high level of stress (*e.g.*, high levels of temperature, voltage, pressure, vibration, cycling rate, or load) in order to provoke early failures. These methods are called the accelerated life tests. The results of this test are then used to estimate the life distribution of the product.

Engineers in manufacturing industries have used accelerated life test (ALT) experiments for many decades. The purpose of ALT experiments is to acquire reliable information quickly.

According to Bai *et al.* [4] and Nelson [17], one way of applying stress to the test is a step-stress scheme which allows the stress setting of a unit to be changed at pre-specified times or upon the occurrence of a fixed number of failures. This scheme is called step stress accelerated life test (SSALT), which is considered in this paper.

To implement the SSALT, we first apply a low stress to all products, if a product endures the stress (does not fail) we apply a higher stress, if only one change of the stress level is done, it is called a simple step-stress accelerated life test.The objective of the SSALT experiment is to estimate the percentile life or reliability prediction by choosing the optimal time of increasing the level of stress that leads to the most accurate estimate. Our main objective is to choose the times to change the stress level in such a way that the variance of estimator of above parameters is minimized under a natural stress level.

The step-stress procedure was first introduced, with the cumulative exposure model, by Nelson [1]. Miller and Nelson [13] provided the optimum simple stress plans for the accelerated life testing, where life products are assumed to have exponentially distributed lifetimes, Bai *et al.* [4] extended the results of Miller and Nelson [13] to the case of censoring. Khamis and Higgins [6, 7] obtained the stress change time which minimizes the asymptotic variance of maximum likelihood estimate of the log mean life at the design condition. Alhadeed and Yang [2] discussed the optimal simple step-stress plan for the Khamis–Higgins model. Most of the available literature on step-stress accelerated life testing deals with the exponential, and Weibull distributions.

The extreme value distribution becomes increasingly important in engineering statistics as a suitable model to represent the phenomena with large extreme observations. In engineering, this distribution is often called the Fréchet model. It is one of the pioneers in extreme value statistics. The Fréchet distribution is one of the probability distributions used to model the maxima extreme events. Thus, the Fréchet distribution is well suited to characterize the random variables of large features and components with a high reliability products. Therefore, it is an important distribution for modeling the statistical behavior of material properties for a variety of engineering applications.

Fréchet distribution is a popular model for lifetimes. Some recent applications have involved the modeling of failure times of air-conditioning systems in jet planes [11] and the modeling of the behavior of off-site AC power failure recovery times at three nuclear plant sites [3] Some results for beta Fréchet distribution are given by [5].

In spite of its popularity, Fréchet distribution has not been used as a lifetime distribution in simple step stress accelerated life test analysis. This paper is the first attempt in this regard. We implement the SSALT analysis and design, by assuming that the failure time of test products follows the Fréchet distribution.

The contents of this paper are organized as follows. The model and basic assumptions are presented in section 2. The maximum likelihood estimators (MLEs) and Fisher information matrix are given in section 3. The optimal test design is derived in section 4, which is followed by a simulation study.

2. MODEL AND TEST PROCEDURE

The Fréchet distribution is a special case of the generalized extreme value distribution. The Fréchet distribution has applications ranging from an accelerated life testing through to earthquakes, floods, horse racing, rainfall, queues in supermarkets, sea currents, wind speeds and track race records. Kotz and Nadarajah [8] give some applications in their book.

To develop appropriate probabilistic models and assess the risks caused by these events, business analysts and engineers frequently use extreme value distributions.

The Fréchet distribution was named after the French mathematician Maurice Fréchet (1878–1973). It is also known as the Extreme Value Type II distribution. It has the cumulative distribution function (CDF) specified by

$$F(t) = \exp\left\{-\left(\frac{t}{\theta}\right)^{-\alpha}\right\} \tag{2.1}$$

for $t > 0$, $\alpha > 0$ and $\theta > 0$. The corresponding probability density function (PDF) is

$$f(t) = \frac{\alpha}{\theta}\left(\frac{t}{\theta}\right)^{-\alpha-1}\exp\left\{-\left(\frac{t}{\theta}\right)^{-\alpha}\right\},$$

where α is a shape parameter and θ is a scale parameter. In engineering applications shape parameter is usually greater than 2.

In a simple SSALT, all n products are initially placed on the test at a lower stress level S_1, and run until time τ when the stress is changed to S_2. The test is continued until all the products run to failure or until a predetermined censoring time T, whichever occurs first. S_0 is stress level at a typical operating condition. Such a test is called a simple step-stress test because it uses only two stress levels. Total n_i failures are observed at time t_{ij}, $j = 1, 2, ..., n_i$, while testing at stress level S_i, $i = 1, 2$, and $n_c = n - n_1 - n_2$ products remain unfailed and censored at time T.

2.1. Basic assumptions

The basic assumptions are:

1. Two stress levels S_1 and S_2 ($S_1 < S_2$) are used in the test.
2. For any level of stress, the life distribution of the test product follows a Fréchet distribution with the CDF (2.1).
3. The scale parameter θ_i at stress level i, $i = 0, 1, 2$ is a log-linear function of stress, *i.e.*,
$$\log(\theta_i) = \beta_0 + \beta_1 S_i$$
for $i = 0, 1, 2$, where β_0 and β_1 are unknown parameters depending on the nature of the product, and the method of test.
4. A cumulative exposure model holds, *i.e.*, the remaining life of a test product depends only on the cumulative exposure it has seen [10].
5. The lifetimes of the test products are identically distributed random variables.

From these assumptions, the CDF of a test product under simple step-stress test is

$$G(t) = \begin{cases} \exp\left\{-\left(\dfrac{t}{\theta_1}\right)^{-\alpha}\right\}, & 0 \le t < \tau\,, \\ \exp\left\{-\left(\dfrac{\tau}{\theta_1}+\dfrac{t-\tau}{\theta_2}\right)^{-\alpha}\right\}, & \tau \le t < \infty\,. \end{cases} \tag{2.2}$$

The corresponding PDF is

$$g(t) = \begin{cases} \dfrac{\alpha}{\theta_1}\left(\dfrac{t}{\theta_1}\right)^{-\alpha-1}\exp\left\{-\left(\dfrac{t}{\theta_1}\right)^{-\alpha}\right\}, & 0 \le t < \tau\,, \\ \dfrac{\alpha}{\theta_2}\left(\dfrac{\tau}{\theta_1}+\dfrac{t-\tau}{\theta_2}\right)^{-\alpha-1}\exp\left\{-\left(\dfrac{\tau}{\theta_1}+\dfrac{t-\tau}{\theta_2}\right)^{-\alpha}\right\}, & \tau \le t < \infty\,. \end{cases}$$

3. MAXIMUM LIKELIHOOD ESTIMATORS

The likelihood function under type I censoring can be written as

$$L(\theta_1,\theta_2,\alpha;t) = \prod_{j=1}^{n_1} g(t_{1j}) \prod_{j=1}^{n_2} g(t_{2j}) \left[1-G(T)\right]^{n_c}.$$

Therefore,

$$\begin{aligned} L(\theta_1,\theta_2,\alpha;t) &= \alpha^{n_1+n_2}\left(\frac{1}{\theta_1}\right)^{n_1}\left(\frac{1}{\theta_2}\right)^{n_2}\prod_{j=1}^{n_1}\left(\frac{t_{1j}}{\theta_1}\right)^{-\alpha-1}\exp\left\{-\sum_{j=1}^{n_1}\left(\frac{t_{1j}}{\theta_1}\right)^{-\alpha}\right\} \\ &\quad\cdot\prod_{j=1}^{n_2}\left(\frac{\tau}{\theta_1}+\frac{t_{2j}-\tau}{\theta_2}\right)^{-\alpha-1}\exp\left\{-\sum_{j=1}^{n_2}\left(\frac{\tau}{\theta_1}+\frac{t_{2j}-\tau}{\theta_2}\right)^{-\alpha}\right\} \\ &\quad\cdot\left(1-\exp\left\{-\left(\frac{\tau}{\theta_1}+\frac{T-\tau}{\theta_2}\right)^{-\alpha}\right\}\right)^{n_c}. \end{aligned}$$

It is usually easier to maximize the logarithm of the likelihood function rather than the likelihood function itself. The logarithm of the likelihood function is

$$\begin{aligned} \ell &= \log L(\theta_1,\theta_2,\alpha;t) \\ &= (n_1+n_2)\log\alpha - n_1\log\theta_1 - n_2\log\theta_2 \\ &\quad-(\alpha+1)\sum_{j=1}^{n_1}\log\left(\frac{t_{1j}}{\theta_1}\right)-\sum_{j=1}^{n_1}\left(\frac{t_{1j}}{\theta_1}\right)^{-\alpha} \\ &\quad-(\alpha+1)\sum_{j=1}^{n_2}\log\left(\frac{\tau}{\theta_1}+\frac{t_{2j}-\tau}{\theta_2}\right)-\sum_{j=1}^{n_2}\left(\frac{\tau}{\theta_1}+\frac{t_{2j}-\tau}{\theta_2}\right)^{-\alpha} \\ &\quad+n_c\log\left(1-\exp\left\{-\left(\frac{\tau}{\theta_1}+\frac{T-\tau}{\theta_2}\right)^{-\alpha}\right\}\right). \end{aligned} \tag{3.1}$$

If at least one failure occurred before τ and T, MLEs of θ_1 and θ_2 do exist. In this case, MLEs of θ_1, θ_2 and α and hence the MLEs of β_0 and β_1 by the invariance property, they can be obtained through setting to zero the first partial derivatives of the log likelihood function with respect to θ_1, θ_2 and α. The system of equations is:

$$\begin{aligned}\frac{\partial \ell}{\partial \theta_1} &= \alpha\frac{n_1}{\theta_1} - \sum_{j=1}^{n_1}\frac{\alpha}{\theta_1}A_j^{-\alpha} + (\alpha+1)\sum_{j=1}^{n_2}\frac{\tau}{\theta_1^2}B_j^{-1}\\ &\quad -\frac{\alpha\tau}{\theta_1^2}\sum_{j=1}^{n_2}B_j^{-\alpha-1} - \frac{\alpha\, n_c \tau\,\theta_2}{\theta_1 E}C^{-\alpha}D^{-1} \;=\; 0\,,\end{aligned} \tag{3.2}$$

$$\begin{aligned}\frac{\partial \ell}{\partial \theta_2} &= -\frac{n_2}{\theta_2} + (\alpha+1)\sum_{j=1}^{n_2}\frac{t_{2j}-\tau}{\theta_2^2}B_j^{-1} - \alpha\sum_{j=1}^{n_2}\frac{t_{2j}-\tau}{\theta_2^2}B_j^{-\alpha-1}\\ &\quad -\frac{\alpha\, n_c (T-\tau)\,\theta_1}{\theta_2 DE}C^{-\alpha} \;=\; 0\,,\end{aligned} \tag{3.3}$$

$$\begin{aligned}\frac{\partial \ell}{\partial \alpha} &= \frac{n_1+n_2}{\alpha} - \sum_{j=1}^{n_1}\log A_j + \sum_{j=1}^{n_1}A_j^{-\alpha}\log A_j - \sum_{j=1}^{n_2}\log B_j\\ &\quad +\sum_{j=1}^{n_2}B_j^{-\alpha}\log B_j + n_c C^{-\alpha}D^{-1}\log C \;=\; 0\,,\end{aligned} \tag{3.4}$$

where $A_j = \frac{t_{1j}}{\theta_1}$, $j=1,2,...,n_1$, $B_j = \frac{\tau}{\theta_1} + \frac{t_{2j}-\tau}{\theta_2}$, $j=1,2,...,n_2$, $C = \frac{\tau}{\theta_1} + \frac{T-\tau}{\theta_2}$, $D = 1-\exp\{C^{-\alpha}\}$ and $E = \theta_1(T-\tau) + \theta_2\tau$.

Given that, it is difficult to obtain a closed form solution to the nonlinear equations (3.2), (3.3) and (3.4), a numerical method is used to solve these equations. By solving these equations, the MLEs $(\theta_1, \theta_2, \alpha)$ and hence MLEs (β_0, β_1) can be obtained.

We have used from optimization tool in Matlab software for finding a maximum of a function of several variables.

The Fisher information essentially describes the amount of information data provide about an unknown parameter. It has applications in finding the variance of an estimator, as well as in the asymptotic behavior of maximum likelihood estimates. The inverse of the Fisher information matrix is an estimator of the asymptotic covariance matrix.

The Fisher information matrix $F(\theta_1, \theta_2, \alpha)$ is obtained through taking expectation on the negative of the second partial derivatives of $\ell(\theta_1, \theta_2, \alpha)$ with respect to θ_1, θ_2 and α.

$$F = n\begin{bmatrix} A_{11} & A_{12} & A_{13}\\ A_{12} & A_{22} & A_{23}\\ A_{13} & A_{23} & A_{33}\end{bmatrix}.$$

The calculation detail is presented in Appendix I and II. Therefore, the elements of F are given as

$$
\begin{aligned}
A_{11} &= \mathrm{E}\Big[-\frac{1}{n}\cdot\frac{\partial^2\ell}{\partial\theta_1^2}\Big] = \frac{\alpha}{\theta_1^2}\,e_1 + \frac{\alpha(\alpha-1)}{\theta_1^2}\,I_1 + (\alpha+1)\,I_2 + I_3 \\
&\quad - \alpha\tau\,(T-\tau)\,\frac{\theta_2\,C^{-\alpha}}{\theta_1 D E^2}\,e_c + \alpha^2\tau^2\,\frac{\theta_2\,C^{-1-2\alpha}(1-D)}{\theta_1^3 E D^2}\,e_c \\
&\quad + \alpha^2\tau^2\,\frac{\theta_2\,C^{-1-\alpha}}{\theta_1^3 E D}\,e_c - \alpha\tau\,\frac{\theta_2\,C^{-\alpha}}{\theta_1^2 E D}\,e_c\,,
\end{aligned}
$$

$$
\begin{aligned}
A_{22} &= \mathrm{E}\Big[-\frac{1}{n}\cdot\frac{\partial^2\ell}{\partial\theta_2^2}\Big] = -\frac{e_2}{\theta_2^2} + (\alpha+1)\,I_4 + I_5 + \frac{\alpha^2(T-\tau)^2\,\theta_1 C^{1-2\alpha}(1-D)}{\theta_2^3\,E D^2}\,e_c \\
&\quad - \frac{\alpha\tau\,(T-\tau)\,\theta_1 C^{-\alpha}}{\theta_2\,E^2 D}\,e_c + \frac{\alpha^2(T-\tau)^2\,\theta_1 C^{-1-\alpha}}{\theta_2^3\,E D}\,e_c \\
&\quad - \frac{\alpha\,\theta_1(T-\tau)\,C^{-\alpha}}{\theta_2^2\,E D}\,e_c\,,
\end{aligned}
$$

$$
\begin{aligned}
A_{33} &= \mathrm{E}\Big[-\frac{1}{n}\cdot\frac{\partial^2\ell}{\partial\alpha^2}\Big] = \frac{1}{\alpha^2}\,e_1 + e_2 + I_6 + \frac{C^{-2\alpha}\,(\log C)^2\,(1-D)}{D^2}\,e_c \\
&\quad + \frac{C^{-\alpha}\,(\log C)^2}{D^2}\,e_c\,,
\end{aligned}
$$

$$
\begin{aligned}
A_{12} &= \mathrm{E}\Big[-\frac{1}{n}\cdot\frac{\partial^2\ell}{\partial\theta_1\partial\theta_2}\Big] = -\frac{\tau(\alpha+1)}{\theta_1^2}\,I_7 + \frac{\alpha(\alpha+1)\tau}{\theta_1^2}\,I_8 \\
&\quad + \alpha^2\tau\,(T-\tau)\,\frac{C^{-1-2\alpha}(1-D)}{\theta_1\theta_2\,E D^2}\,e_c - \alpha\tau^2\,\frac{\theta_2\,C^{-\alpha}}{\theta_1 E^2 D}\,e_c \\
&\quad + \alpha^2\tau\,(T-\tau)\,\frac{C^{-\alpha-1}}{\theta_1\theta_2\,E D}\,e_c + \alpha\tau\,\frac{C^{-\alpha}}{\theta_1 E D}\,e_c\,,
\end{aligned}
$$

$$
\begin{aligned}
A_{13} &= \mathrm{E}\Big[-\frac{1}{n}\cdot\frac{\partial^2\ell}{\partial\theta_1\partial\alpha}\Big] = -\frac{e_1}{\theta_1} + \frac{1}{\theta_1}\,I_9 + \frac{\tau}{\theta_1^2}\,I_{10} + \frac{\tau}{\theta_1^2}\,I_{11} + \tau\,\theta_2\,\frac{C^{-\alpha}}{\theta_1 E D}\,e_c \\
&\quad - \alpha\tau\theta_2\,\frac{C^{-\alpha}\log C}{\theta_1 E D}\,e_c - \alpha\tau\theta_2\,\frac{C^{-2\alpha}(1-D)\,\log C}{\theta_1 E D^2}\,e_c\,,
\end{aligned}
$$

$$
\begin{aligned}
A_{23} &= \mathrm{E}\Big[-\frac{1}{n}\cdot\frac{\partial^2\ell}{\partial\theta_2\partial\alpha}\Big] = I_{12} + I_{13} - \alpha\,(T-\tau)\,\theta_1\,\frac{C^{-2\alpha}(1-D)\,\log C}{\theta_2\,E D^2}\,e_c \\
&\quad + \theta_1(T-\tau)\,\frac{C^{-\alpha}}{\theta_2\,E D}\,e_c - \alpha\,(T-\tau)\,\theta_1\,\frac{C^{\alpha}\log C}{\theta_2\,E D}\,e_c\,.
\end{aligned}
$$

where the detailed calculation for I_1 to I_{13}, and e_1, e_2 and e_c in the formulas above are in Appendix I and Appendix II, respectively.

The asymptotic variance of the desired estimates is then obtained using the above Fisher information matrix, which leads to the optimization criteria.

4. OPTIMUM TEST DESIGN

As mentioned earlier, for the purpose of optimization, two criteria are considered. The first criterion (Criterion I) is minimizing the asymptotic variance (AV) of the MLE of the logarithm of the percentile life under usual operating conditions, which is used when the percentile life is the desired estimate. Furthermore, we can minimize the AV of reliability estimate at time ξ under usual operating conditions. We call this criterion as the second criterion (Criterion II) and is used when we want to predict reliability.

We will show the optimal hold times achieved by criterion I and II, with the symbols of τ^* and τ^+, respectively.

4.1. Criterion I

As mentioned above, in this criterion, we try to minimize the AV of the MLE of the logarithm of percentile life under the usual operating conditions. This is the most commonly used criterion.

The reliability function at time t under the usual operating condition, S_0, is:

$$R_0(t) \;=\; 1 - G_0(t) \;=\; 1 - \exp\left\{-\left(\frac{t}{\theta_0}\right)^{-\alpha}\right\}.$$

For a specified reliability R, the $100\,(1-R)$-th percentile life under the usual operating condition, S_0, is:

$$t_R \;=\; \theta_0\Big(-\log\left(1-R\right)\Big)^{-1/\alpha}.$$

From assumption 3 and the definition, $x = \frac{S_1 - S_0}{S_2 - S_0}$, we obtain $S_0 = \frac{S_1 - x S_2}{1-x}$, thus,

$$\log\theta_0 \;=\; \frac{\log\theta_1 - x\log\theta_2}{1-x}\,. \tag{4.1}$$

Therefore, the MLE of the log of the $100\,(1-R)$-th percentile life of the Fréchet distribution with a specified reliability R under the usual operating condition, S_0, is:

$$\begin{aligned}\log\left(\widehat{t}_R\right) \;&=\; \log\widehat{\theta}_0 - \frac{1}{\widehat{\alpha}}\log\Big(-\log\left(1-R\right)\Big)\\ &=\; \frac{\log\widehat{\theta}_1 - x\log\widehat{\theta}_2}{1-x} - \frac{\log\Big(-\log\left(1-R\right)\Big)}{\widehat{\alpha}}\,.\end{aligned}$$

The optimality criterion used for the SSALT design is to minimize the AV of the MLE of the log of the $100\,(1-R)$-th percentile life of the Fréchet distribution at S_0 with a specified reliability R. When $R = 0.5$, $\log(\widehat{t}_R)$ is the logarithm of the median life at usual operating conditions with stress level S_0. To obtain the $AV[\log(\widehat{t}_R)]$, we use the delta method which described in Appendix III.

The optimal hold time τ_0^* at which $AV[\log(\widehat{t}_R)]$ reaches its minimum value leads to the optimal plan:

$$AV\Big[\log\big(\widehat{t}_R\big)\Big] \;=\; AV\left[\frac{\log\widehat{\theta}_1 - x\log\widehat{\theta}_2}{1-x} - \frac{\log\Big(-\log\,(1-R)\Big)}{\widehat{\alpha}}\right] \;=\; H_1\,\widehat{F}^{-1}H_1'\;,$$

where $\widehat{F}$ is estimated the Fisher information matrix and H_1 is the row vector of the first derivative of $\log(\widehat{t}_R)$ with respect to $\widehat{\theta}_1$, $\widehat{\theta}_2$ and $\widehat{\alpha}$; and in practice, the values of $(\widehat{\theta}_1, \widehat{\theta}_2, \widehat{\alpha})$ are obtained from a previous experience based on a similar data, or based on a preliminary test result.

$$H_1 \;=\; \left[\frac{1}{\widehat{\theta}_1(1-x)},\, \frac{x}{\widehat{\theta}_2\,(x-1)},\, \frac{\log\Big(-\log\,(1-R)\Big)}{\widehat{\alpha}^2}\right].$$

4.2. Criterion II

Reliability prediction is an important factor in a product design and during the developmental testing process. In order to accurately estimate the product reliability, the test design criterion is defined to minimize the AV of the reliability estimate at a time ξ under the normal operating conditions.

The MLE of reliability at ξ from the Fréchet distribution at the usual operating stress level, S_0, is:

$$\widehat{R}_{S_0}(\xi) \;=\; 1-\exp\left\{-\left(\frac{\xi}{\widehat{\theta}_0}\right)^{-\widehat{\alpha}}\right\} \;=\; 1-\exp\left\{-\,\exp\left\{-\,\widehat{\alpha}\log\xi + \widehat{\alpha}\log\widehat{\theta}_0\right\}\right\},$$

where, by using (4.1), we have

$$\widehat{R}_{S_0}(\xi) \;=\; 1-\exp\left\{-\,\exp\left\{-\,\widehat{\alpha}\log\xi + \widehat{\alpha}\,\frac{\log\widehat{\theta}_1 - x\log\widehat{\theta}_2}{1-x}\right\}\right\}.$$

The AV of the reliability estimate at time ξ under normal operating conditions, by using the delta method, can be obtained as:

$$AV\Big[\widehat{R}_{S_0}(\xi)\Big] \;=\; AV\left[1-\exp\left\{-\,\exp\left\{-\,\widehat{\alpha}\log\xi + \widehat{\alpha}\log\widehat{\theta}_0\right\}\right\}\right] \;=\; H_2\,\widehat{F}^{-1}H_2'\;, \tag{4.2}$$

where H_2 is the row vector of the first derivative of $\widehat{R}_{S_0}(\xi)$ with respect to $\widehat{\theta}_1$, $\widehat{\theta}_2$ and $\widehat{\alpha}$, *i.e.*, $H_2 = \left[H_{11}, H_{12}, H_{13}\right]$, where its components are given below. In practice, Based on experience, some historical data or a preliminary test can be used to get the values of $(\widehat{\theta}_1, \widehat{\theta}_2, \widehat{\alpha})$.

$$\begin{aligned}
H_{11} &= \frac{\widehat{\alpha}\,\xi^{-\widehat{\alpha}}}{\widehat{\theta}_1(1-x)} \exp\left\{-\,\xi^{-\widehat{\alpha}}\, e^{\widehat{\alpha}\frac{\log\widehat{\theta}_1 - x\log\widehat{\theta}_2}{1-x}} + \widehat{\alpha}\,\frac{\log\widehat{\theta}_1 - x\log\widehat{\theta}_2}{1-x}\right\},\\
H_{12} &= \frac{x\,\widehat{\alpha}\,\xi^{-\widehat{\alpha}}}{\widehat{\theta}_2\,(x-1)} \exp\left\{-\,\xi^{-\widehat{\alpha}}\, e^{\widehat{\alpha}\frac{\log\widehat{\theta}_1 - x\log\widehat{\theta}_2}{1-x}} + \widehat{\alpha}\,\frac{\log\widehat{\theta}_1 - x\log\widehat{\theta}_2}{1-x}\right\},\\
H_{13} &= \frac{1}{x-1}\left(\exp\left\{-\,\xi^{-\widehat{\alpha}}\, e^{\widehat{\alpha}\frac{\log\widehat{\theta}_1 - x\log\widehat{\theta}_2}{1-x}} + \widehat{\alpha}\,\frac{\log\widehat{\theta}_1 - x\log\widehat{\theta}_2}{1-x}\right\}\right.\\
&\quad \left.\cdot\, \xi^{-\widehat{\alpha}}\left(\log\xi - x\log\xi - \log\widehat{\theta}_1 + x\log\widehat{\theta}_2\right)\right).
\end{aligned}$$

The value τ_0^+ that minimizes $AV\left[\widehat{R}_{S_0}(\xi)\right]$, given by equation (4.2), leads to the optimal SSALT plan.

4.3. Simulation study

The main objective of this simulation study is numerical investigation for illustrating the theoretical results of both estimation and optimal design problems given in this paper. Considering type I censoring, data were generated from Fréchet distribution under SSALT for different combinations of the true parameter values of θ_1, θ_2 and α. The true parameters values used here are $(1.5, 1, 1)$ and $(2.5, 2, 1.5)$. In addition, $\tau = 2.5$ and $T = 5$ have been considered. The samples sizes considered are $n = 100, 200, 300, 400, 500, 1000$ each with ten thousand replications. A numerical method is used for the MLEs of θ_1, θ_2 and α. The nonlinear likelihood equations, (3.2), (3.3) and (3.4), were solved iteratively. The MLEs, their mean square errors (MSEs) and their relative errors (REs) are reported in Table 1 for different sample sizes and different true values of the parameters. The results provide insight into the sampling behavior of the estimators. They indicate that the MLEs approximate the true values of the parameters as the sample size n increases. Similarly, the MSEs and REs decrease with increasing the sample size.

To illustrate the procedure of the optimum test design, we proposed a standardized model. A standardized censoring time $T_0 = 1$ is assumed, and the standardized scale parameter $\eta_i = \frac{\theta_i}{T}$ is defined. The standardized hold time τ_0 is also defined as the ratio of the hold time to the censoring time $\tau_0 = \frac{\tau}{T}$. Thus the value of τ_0 that minimizes AV is the optimal standardized hold time, and the optimal hold time is derived from $\tau^* = \tau_0^* \cdot T$ and $\tau^+ = \tau_0^+ \cdot T$, with respect to criterion I and II. Using the standardized model, we eliminate the input value of censoring time and embed it in the standardized scale parameters.

Table 1: The MLEs of the parameters, and the associated MSE and RE for different sample sizes.

n	Parameter	$(\theta_1=1.5,\ \theta_2=1,\ \alpha=1)$			$(\theta_1=2.5,\ \theta_2=2,\ \alpha=1.5)$		
		Estimate	MSE	RE	Estimate	MSE	RE
$n=100$	θ_1	1.4404	0.0108	0.0397	2.5228	0.0500	0.0091
	θ_2	1.0381	0.0771	0.0381	2.1101	0.2825	0.0091
	α	1.0307	0.0103	0.0307	1.5362	0.0367	0.0241
$n=200$	θ_1	1.4563	0.0058	0.0291	2.5086	0.0238	0.0034
	θ_2	1.0273	0.0461	0.0273	2.0541	0.1372	0.0271
	α	1.0205	0.0050	0.0205	1.5180	0.0186	0.0120
$n=300$	θ_1	1.4640	0.0040	0.0240	2.5061	0.0156	0.0024
	θ_2	1.0196	0.0317	0.0196	2.0374	0.0877	0.0187
	α	1.0160	0.0033	0.0160	1.5128	0.0121	0.0085
$n=400$	θ_1	1.4682	0.0031	0.0212	2.5044	0.0118	0.0018
	θ_2	1.0142	0.0237	0.0142	2.0279	0.0650	0.0140
	α	1.0133	0.0025	0.0133	1.5094	0.0090	0.0063
$n=500$	θ_1	1.4703	0.0026	0.0198	2.5036	0.0095	0.0014
	θ_2	1.0132	0.0194	0.0132	2.0200	0.0501	0.0100
	α	1.0177	0.0020	0.0117	1.5073	0.0071	0.0049
$n=1000$	θ_1	1.4801	0.0012	0.0132	2.5021	0.0046	8.4513×10^{-4}
	θ_2	1.0055	0.0094	0.0055	2.0106	0.0249	0.0053
	α	1.0073	9.5334×10^{-4}	0.0073	1.5034	0.0036	0.0023

Now, the numerical examples are given for calculating the optimal standardized hold times of the simple SSALT under both criteria.

In the first example, we suppose that a simple SSALT to estimate the percentile life of the Fréchet distribution under the usual operating condition with a specified reliability R. For the given values of $\theta_1 = 900$, $\theta_2 = 400$, $\alpha = 2$, $T = 1000$, $x = 0.5$ and assuming $R = 0.5$, we determine the optimal hold time τ^*. Based on the above transformation, the standardized parameters are obtained as $\eta_1 = 0.9$ and $\eta_2 = 0.4$. Using the criterion I, the optimal standardized hold time is obtained $\tau_0^* = 0.8165$. So, the optimum stress change time is obtained $\tau^* = 816.5$.

Sensitivity analysis is performed to examine the effect of the changes in the pre-estimated parameters $(\theta_1, \theta_2, \alpha)$ on the optimal hold time τ. Its objective is to identify the sensitive parameters, which need to be estimated with special care to minimize the risk of obtaining an erroneous optimal solution. According to the definition of x and R; and since they take different values, we also examine the impact of changes in their values.

Table 2 presents the standardized optimal hold time for the specified values of $n = 30$, $R = 0.5$, $x = 0.5$, $\alpha = 2$, $\eta_1 = 0.3, 0.5, ..., 1.7$ and $\eta_2 = 0.1, 0.3, ..., 1.5$. From this table, we can see that as η_1 increases, the optimal standardized stress change time slightly increases. And also, as η_2 increases, then slightly decreases.

Table 2: Optimal standardized hold time τ_0^* and AV versus changes in η_1 and η_2 with criterion I ($\alpha=2$, $x=0.5$, $R=0.5$).

η_1	η_2							
	0.1	0.3	0.5	0.7	0.9	1.1	1.3	1.5
0.3	$\tau_0^*=0.3625$ AV = 8.9925							
0.5	$\tau_0^*=0.6085$ AV = 9.0273	$\tau_0^*=0.5985$ AV = 9.0828						
0.7	$\tau_0^*=0.8385$ AV = 9.0832	$\tau_0^*=0.7635$ AV = 9.4182	$\tau_0^*=0.7165$ AV = 9.8620					
0.9	$\tau_0^*=0.9175$ AV = 9.6894	$\tau_0^*=0.8435$ AV = 10.3295	$\tau_0^*=0.7925$ AV = 10.8826	$\tau_0^*=0.7475$ AV = 11.3462				
1.1	$\tau_0^*=0.9425$ AV = 10.9082	$\tau_0^*=0.8725$ AV = 11.4226	$\tau_0^*=0.8215$ AV = 11.8557	$\tau_0^*=0.7865$ AV = 12.2794	$\tau_0^*=0.7625$ AV = 12.7426			
1.3	$\tau_0^*=0.9495$ AV = 11.9406	$\tau_0^*=0.8965$ AV = 12.4060	$\tau_0^*=0.8605$ AV = 13.0544	$\tau_0^*=0.8355$ AV = 13.8146	$\tau_0^*=0.8165$ AV = 14.6650	$\tau_0^*=0.8025$ AV = 15.5903		
1.5	$\tau_0^*=0.9495$ AV = 13.4693	$\tau_0^*=0.9195$ AV = 14.3669	$\tau_0^*=0.8915$ AV = 15.6970	$\tau_0^*=0.8705$ AV = 17.1575	$\tau_0^*=0.8545$ AV = 18.7231	$\tau_0^*=0.8425$ AV = 20.3801	$\tau_0^*=0.8325$ AV = 22.1190	
1.7	$\tau_0^*=0.9495$ AV = 17.3690	$\tau_0^*=0.9355$ AV = 18.7619	$\tau_0^*=0.9115$ AV = 21.3211	$\tau_0^*=0.8935$ AV = 24.0302	$\tau_0^*=0.8805$ AV = 26.8841	$\tau_0^*=0.8695$ AV = 29.8769	$\tau_0^*=0.8605$ AV = 33.0031	$\tau_0^*=0.8535$ AV = 36.2555

Figure 1 shows the sensitivity of the initially estimated parameters with respect to the criterion I. We can see:

1. The optimal value of τ^*, slightly increases as η_1 and α increase for smaller values of η_1 and α, and converges for larger values of η_1 and α;

2. The optimal value of τ^*, slightly decreases as η_2, R and x increase, and it is not too sensitive to parameters η_2 and x.

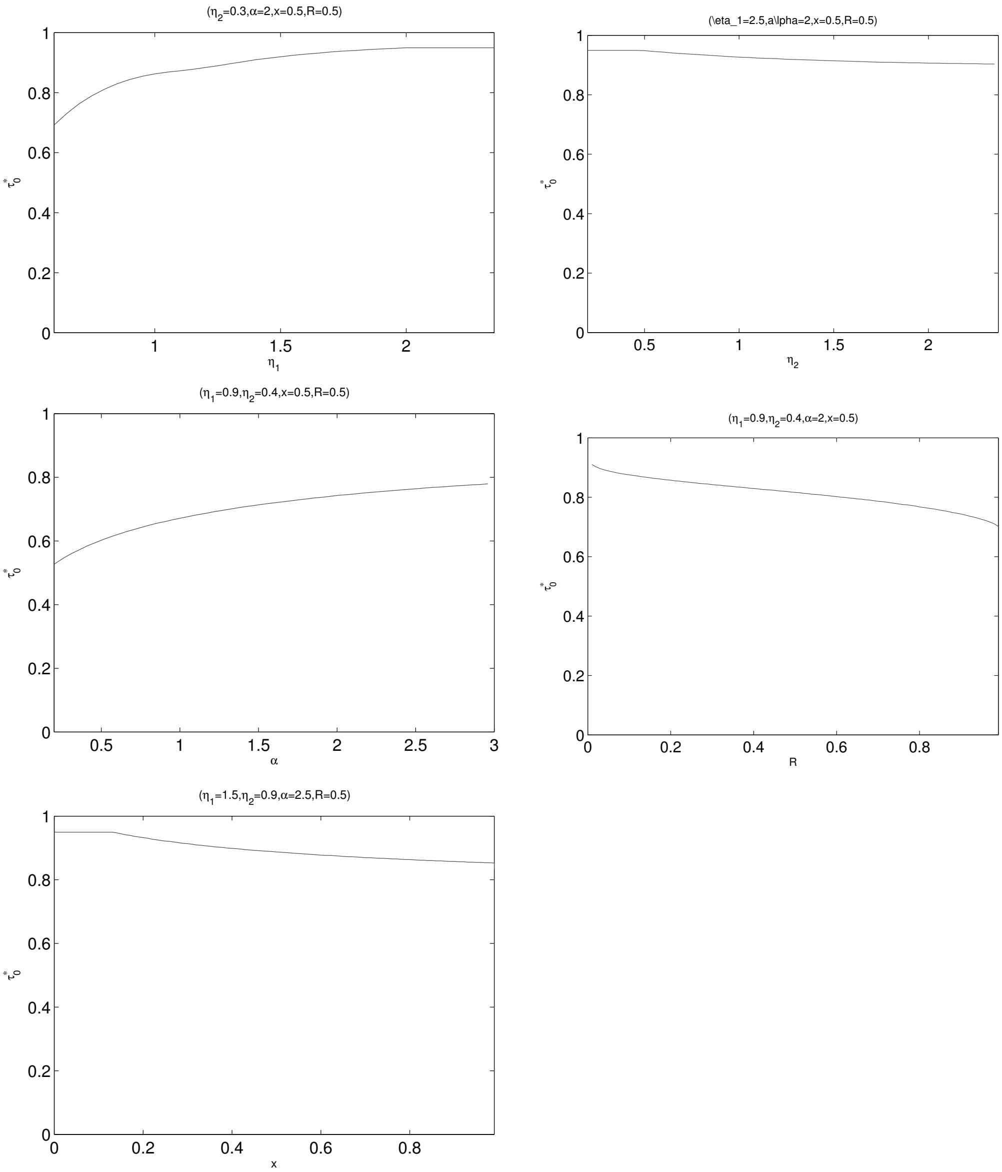


Figure 1: Optimal standardized hold time versus changes in initial parameters under criterion I.

Table 3: Optimal standardized hold time τ_0^+ and AV versus changes in η_1 and η_2 with criterion II ($\alpha = 2$, $x = 0.5$, $\xi = 10$).

η_1	η_2							
	0.1	0.3	0.5	0.7	0.9	1.1	1.3	1.5
0.3	$\tau_0^+ = 0.5155$ $\mathrm{AV} = 0.0068$							
0.5	$\tau_0^+ = 0.7595$ $\mathrm{AV} = 0.2379$	$\tau_0^+ = 0.7595$ $\mathrm{AV} = 0.0054$						
0.7	$\tau_0^+ = 0.8815$ $\mathrm{AV} = 1.7952$	$\tau_0^+ = 0.8585$ $\mathrm{AV} = 0.0649$	$\tau_0^+ = 0.8415$ $\mathrm{AV} = 0.0116$					
0.9	$\tau_0^+ = 0.9195$ $\mathrm{AV} = 4.5875$	$\tau_0^+ = 0.8995$ $\mathrm{AV} = 0.4237$	$\tau_0^+ = 0.8865$ $\mathrm{AV} = 0.0829$	$\tau_0^+ = 0.8765$ $\mathrm{AV} = 0.0274$				
1.1	$\tau_0^+ = 0.9275$ $\mathrm{AV} = 3.5379$	$\tau_0^+ = 0.9065$ $\mathrm{AV} = 1.5776$	$\tau_0^+ = 0.8875$ $\mathrm{AV} = 0.3703$	$\tau_0^+ = 0.8695$ $\mathrm{AV} = 0.1287$	$\tau_0^+ = 0.8535$ $\mathrm{AV} = 0.0568$			
1.3	$\tau_0^+ = 0.9335$ $\mathrm{AV} = 0.6480$	$\tau_0^+ = 0.9075$ $\mathrm{AV} = 3.6960$	$\tau_0^+ = 0.8845$ $\mathrm{AV} = 1.0889$	$\tau_0^+ = 0.8685$ $\mathrm{AV} = 0.4059$	$\tau_0^+ = 0.8555$ $\mathrm{AV} = 0.1851$	$\tau_0^+ = 0.8455$ $\mathrm{AV} = 0.0970$		
1.5	$\tau_0^+ = 0.9405$ $\mathrm{AV} = 0.0226$	$\tau_0^+ = 0.9225$ $\mathrm{AV} = 6.4165$	$\tau_0^+ = 0.9065$ $\mathrm{AV} = 2.6665$	$\tau_0^+ = 0.8955$ $\mathrm{AV} = 1.1107$	$\tau_0^+ = 0.8865$ $\mathrm{AV} = 0.5355$	$\tau_0^+ = 0.8795$ $\mathrm{AV} = 0.2908$	$\tau_0^+ = 0.8735$ $\mathrm{AV} = 0.1727$	
1.7	$\tau_0^+ = 0.9445$ $\mathrm{AV} = 9.4376 \times 10^{-5}$	$\tau_0^+ = 0.9315$ $\mathrm{AV} = 9.0884$	$\tau_0^+ = 0.9205$ $\mathrm{AV} = 6.1714$	$\tau_0^+ = 0.9115$ $\mathrm{AV} = 2.9891$	$\tau_0^+ = 0.9045$ $\mathrm{AV} = 1.5486$	$\tau_0^+ = 0.8995$ $\mathrm{AV} = 0.8780$	$\tau_0^+ = 0.8945$ $\mathrm{AV} = 0.5370$	$\tau_0^+ = 0.8905$ $\mathrm{AV} = 0.3491$

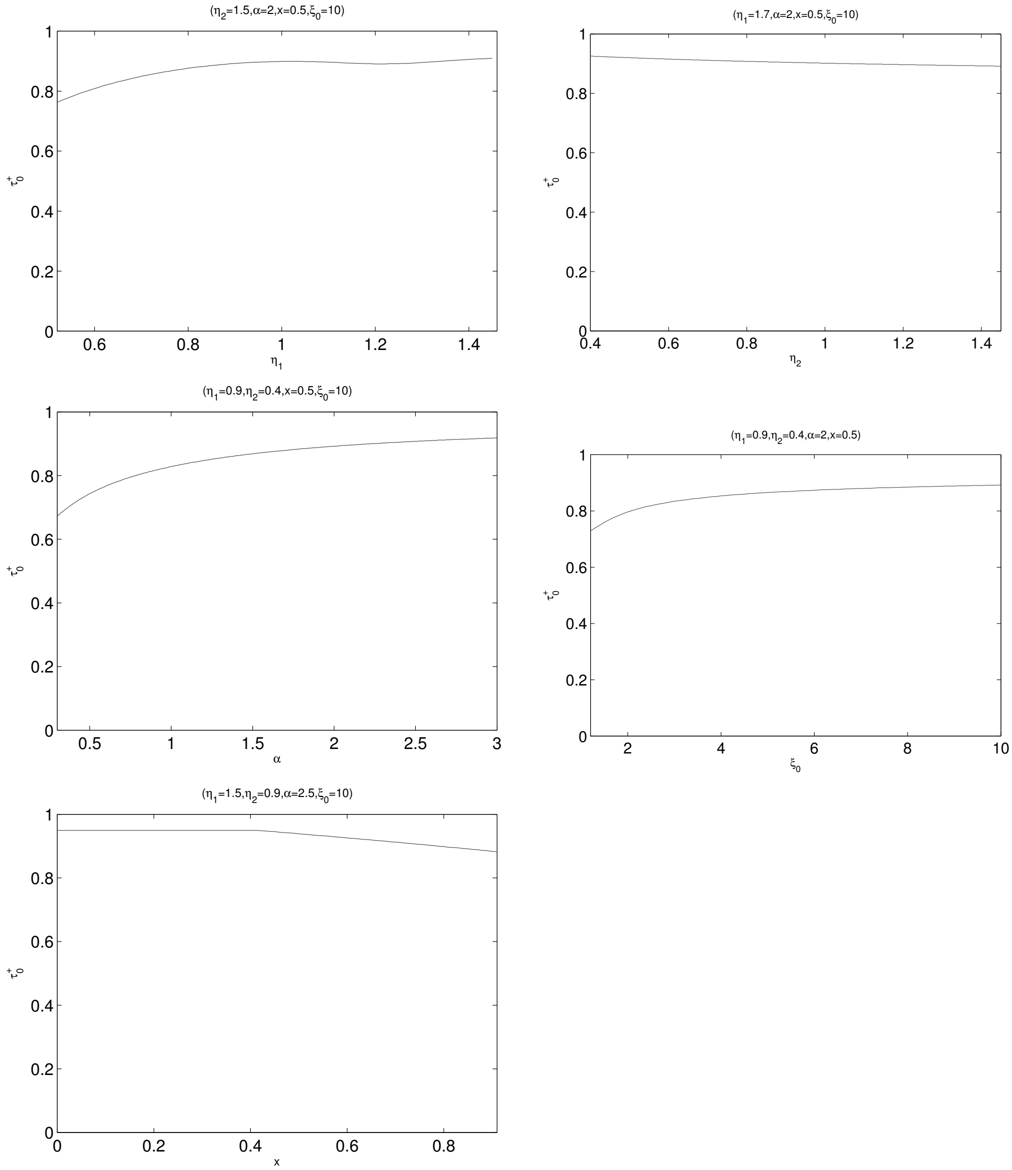


Figure 2: Optimal standardized hold time versus changes in initial parameters under criterion II.

In the second example, we suppose a simple SSALT is run to estimate the reliability at a specified time $\xi = 10000$. The objective is to design a test that achieves the best reliability estimates. To obtain the optimal hold time τ^+, the AV of the reliability estimate at time ξ is minimized. The initial parameters given $\theta_1 = 900,\ \theta_2 = 400,\ \alpha = 2,\ x = 0.5,\ T = 1000$ and $\xi = 10000$. Then the standardized parameters are obtained as $\eta_1 = 0.9,\ \eta_2 = 0.4$ and $\xi_0 = 10$. By criterion II, the optimum standardized hold time is obtained as $\tau_0^+ = 0.8925$ and the optimum stress change time is obtained as $\tau^+ = 8.925$.

Table 3 presents the standardized optimal hold time for the specified values of $n = 30,\ x = 0.5,\ \alpha = 2,\ \xi = 10,\ \eta_1 = 0.3, 0.5, ..., 1.7$ and $\eta_2 = 0.1, 0.3, ..., 1.5$. This table shows that, as η_1 increases, the optimal standardized stress change time slightly increases. And also, as η_2 increases, then slightly decreases.

Figure 2 shows the sensitivity of the initially estimated parameters with respect to criterion II. We can see:

1. The optimal value of τ^+, slightly increases as η_1, α and ξ increase for smaller values of them, and converges for larger values of them;
2. The optimal value of τ^+, slightly decreases as η_2 and x increase, and is not too sensitive to parameters η_2 and x.

5. CONCLUSION

In this paper, we proposed an optimal design of simple step stress accelerated life test with type I censored Fréchet data. Optimizing test plan will lead to an improved parameter estimation which would further lead to a higher quality of inference. The estimation was based on the maximum likelihood.

For the purpose of optimizing, two criteria were considered. These criteria were based on minimizing the AV of the life estimate and the reliability estimate. Furthermore, according to the simulation study, we have found that since the optimal hold times are not too sensitive to the model parameters, thus the proposed design is robust. The results show that the simple SSALT model can be reliably used which would remove the need for examining all the test products and would have economic benefits concerning time and money.

ACKNOWLEDGMENTS

The authors would like to thank the Editor, the Associate Editor and the three referees for carefully reading and for their constructive comments which greatly improved the paper.

APPENDIX I

The Fisher information matrix, can be obtained by taking the expected values of the negative second derivatives with respect to θ_1, θ_2 and α of the function (3.1). The results of these derivatives are given the following:

$$\begin{aligned} I_{11} = -\frac{\partial^2 \ell}{\partial \theta_1^2} &= \frac{\alpha n_1}{\theta_1^2} + \sum_{j=1}^{n_1}\left(\frac{\alpha(\alpha+1)\,A_j^{-\alpha}}{\theta_1^2} - \frac{2\,\alpha\,A_j^{-\alpha}}{\theta_1^2}\right) \\ &+ (\alpha+1)\sum_{j=1}^{n_2}\left(-\frac{\tau^2}{\theta_1^4 B_j^2} - \frac{2\,\tau}{\theta_1^3 B_j}\right) \\ &+ \sum_{j=1}^{n_2}\left(\frac{\alpha\,(\alpha+1)\,\tau^2 B_j^{-2-\alpha}}{\theta_1^4} - \frac{2\,\alpha\,\tau\,B_j^{-1-\alpha}}{\theta_1^3}\right) \\ &- \alpha\tau\,(T-\tau)\,\frac{\theta_2\,C^{-\alpha}}{\theta_1 D E^2}\,n_c + \alpha^2\tau^2\,\frac{\theta_2\,C^{-1-2\alpha}(1-D)}{\theta_1^3 E D^2}\,n_c \\ &+ \alpha^2\tau^2\,\frac{\theta_2\,C^{-1-\alpha}}{\theta_1^3 E D}\,n_c - \alpha\,\tau\,n_c\,\frac{\theta_2\,C^{-\alpha}}{\theta_1^2 E D}\,, \end{aligned}$$

$$\begin{aligned} I_{22} = -\frac{\partial^2 \ell}{\partial \theta_2^2} &= -\frac{n_2}{\theta_2^2} + \alpha\sum_{j=1}^{n_2}\left((\alpha+1)\,F_j^2\,B_j^{-2-\alpha} - \frac{2}{\theta_2}\,B_j^{-\alpha-1}F_j\right) \\ &+ (\alpha+1)\sum_{j=1}^{n_2}\left(-\frac{F_j^2}{\theta_2^2 B_j^2} + \frac{2\,F_j}{\theta_2\,B_j}\right) \\ &+ \frac{\alpha^2(T-\tau)^2\,\theta_1 C^{1-2\alpha}(1-D)}{\theta_2^3\,E D^2}\,n_c - \frac{\alpha\,\tau\,(T-\tau)\,\theta_1 C^{-\alpha}}{\theta_2\,E^2 D}\,n_c \\ &+ \frac{\alpha^2(T-\tau)^2\,\theta_1 C^{-1-\alpha}}{\theta_2^3\,E D}\,n_c - \frac{\alpha\,\theta_1(T-\tau)\,C^{-\alpha}}{\theta_2^2\,E D}\,n_c\,, \end{aligned}$$

$$\begin{aligned} I_{33} = -\frac{\partial^2 \ell}{\partial \alpha^2} &= \frac{n_1+n_2}{\alpha^2} + \sum_{j=1}^{n_2} B_j^{-\alpha}\Big(\log B_j\Big)^2 \\ &+ \frac{C^{-2\alpha}(\log C)^2\,(1-D)}{D^2}\,n_c + \frac{C^{-\alpha}(\log C)^2}{D^2}\,n_c\,, \end{aligned}$$

$$\begin{aligned} I_{12} = -\frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_2} &= -\frac{(\alpha+1)\,\tau}{\theta_1^2}\sum_{j=1}^{n_2}\frac{F_j}{\theta_1^2 B_j^2} + \frac{\alpha\,(\alpha+1)\,\tau}{\theta_1^2}\sum_{j=1}^{n_2} F_j\,B_j^{-2-\alpha} \\ &+ \alpha\,\tau\,\frac{C^{-\alpha}}{\theta_1 E D}\,n_c + \alpha^2\tau\,(T-\tau)\,\frac{C^{-1-2\alpha}(1-D)}{\theta_1\theta_2\,E D^2}\,n_c \\ &- \alpha\,\tau^2\,\frac{\theta_2\,C^{-\alpha}}{\theta_1 E^2 D}\,n_c + \alpha^2\tau\,(T-\tau)\,\frac{C^{-\alpha-1}}{\theta_1\theta_2\,E D}\,n_c\,, \end{aligned}$$

$$
\begin{aligned}
I_{13} = -\frac{\partial^2 \ell}{\partial\theta_1 \partial\alpha} &= -\frac{n_1}{\theta_1} + \frac{1}{\theta_1}\sum_{j=1}^{n_1} A_j^{-\alpha} - \sum_{j=1}^{n_2} \frac{\tau}{\theta_1^2 B_j} \\
&+ \frac{\tau}{\theta_1^2}\sum_{j=1}^{n_2} B_j^{-1-\alpha}\Big(1 - \alpha \log B_j\Big) + \tau\,\theta_2\,\frac{C^{-\alpha}}{\theta_1 ED}\, n_c \\
&- \alpha\tau\theta_2\,\frac{C^{-2\alpha}(1-D)\log C}{\theta_1 ED^2}\, n_c - \alpha\tau\theta_2\,\frac{C^{-\alpha}\log C}{\theta_1 ED}\, n_c\ ,
\end{aligned}
$$

$$
\begin{aligned}
I_{23} = -\frac{\partial^2 \ell}{\partial\theta_2 \partial\alpha} &= \sum_{j=1}^{n_2} \frac{F_j}{B_j} + \sum_{j=1}^{n_2} F_j\, B_j^{-1-\alpha}\Big(1 - \alpha \log B_j\Big) \\
&- \alpha\,(T-\tau)\,\theta_1\,\frac{C^{-2\alpha}(1-D)\log C}{\theta_2\, ED^2}\, n_c \\
&+ \theta_1(T-\tau)\,\frac{C^{-\alpha}}{\theta_2\, ED}\, n_c - \alpha\,(T-\tau)\,\theta_1\,\frac{C^{\alpha}\log C}{\theta_2\, ED}\, n_c\ ,
\end{aligned}
$$

where $F_j = \frac{t_{2j}-\tau}{\theta_2^2}$, $j = 1, 2, ..., n_2$.

The results of the above equations are then used to develop the Fisher information matrix. And also, to simplify the second partial and mixed partial derivatives, the following definitions are made:

$$
I_1 = \mathrm{E}\left[\frac{1}{n}\sum_{j=1}^{n_1} A_j^{-\alpha}\right] = \int_0^{\tau} A_j^{-\alpha} g(t)\, d(t)\ ,
$$

$$
I_2 = \mathrm{E}\left[\frac{1}{n}\sum_{j=1}^{n_2}\left(-\frac{\tau^2}{\theta_1^4 B_j^2} - \frac{2\,\tau}{\theta_1^3 B_j}\right)\right] = \int_{\tau}^{T}\left(-\frac{\tau^2}{\theta_1^4 B_j^2} - \frac{2\,\tau}{\theta_1^3 B_j}\right) g(t)\, d(t)\ ,
$$

$$
\begin{aligned}
I_3 &= \mathrm{E}\left[\frac{1}{n}\sum_{j=1}^{n_2}\left(\frac{\alpha\,(\alpha+1)\,\tau^2\, B_j^{-2-\alpha}}{\theta_1^4} - \frac{2\,\alpha\,\tau\, B_j^{-1-\alpha}}{\theta_1^3}\right)\right] \\
&= \int_{\tau}^{T}\left(\frac{\alpha\,(\alpha+1)\,\tau^2\, B_j^{-2-\alpha}}{\theta_1^4} - \frac{2\,\alpha\,\tau\, B_j^{-1-\alpha}}{\theta_1^3}\right) g(t)\, d(t)\ ,
\end{aligned}
$$

$$
I_4 = \mathrm{E}\left[\frac{1}{n}\sum_{j=1}^{n_2}\left(-\frac{F_j^2}{\theta_2^2 B_j^2} + \frac{2\,F_j}{\theta_2\, B_j}\right)\right] = \int_{\tau}^{T}\left(-\frac{F_j^2}{\theta_2^2 B_j^2} + \frac{2\,F_j}{\theta_2\, B_j}\right) g(t)\, d(t)\ ,
$$

$$
\begin{aligned}
I_5 &= \mathrm{E}\left[\frac{1}{n}\sum_{j=1}^{n_2}\left((\alpha+1)\, F_j^2\, B_j^{-2-\alpha} - \frac{2}{\theta_2}\, B_j^{-\alpha-1} F_j\right)\right] \\
&= \int_{\tau}^{T}\left((\alpha+1)\, F_j^2\, B_j^{-2-\alpha} - \frac{2}{\theta_2}\, B_j^{-\alpha-1} F_j\right) g(t)\, d(t)\ ,
\end{aligned}
$$

$$
I_6 = \mathrm{E}\left[\frac{1}{n}\sum_{j=1}^{n_2} B_j^{-\alpha}\Big(\log B_j\Big)^2\right] = \int_{\tau}^{T} B_j^{-\alpha}\Big(\log B_j\Big)^2 g(t)\, d(t)\ ,
$$

$$
\begin{aligned}
I_7 &= \mathrm{E}\left[\frac{1}{n}\sum_{j=1}^{n_2}\frac{F_j}{\theta_1^2 B_j^2}\right] = \int_\tau^T \frac{F_j}{\theta_1^2 B_j^2}\; g(t)\,d(t)\;, \\
I_8 &= \mathrm{E}\left[\frac{1}{n}\sum_{j=1}^{n_2} F_j\, B_j^{-2-\alpha}\right] = \int_\tau^T F_j\, B_j^{-2-\alpha}\, g(t)\,d(t)\;, \\
I_9 &= \mathrm{E}\left[\frac{1}{n}\sum_{j=1}^{n_1} A_j^{-\alpha}\right] = \int_0^\tau A_j^{-\alpha}\, g(t)\,d(t)\;, \\
I_{10} &= \mathrm{E}\left[\frac{1}{n}\sum_{j=1}^{n_2} B_j^{-1-\alpha}\Big(1-\alpha\log B_j\Big)\right] = \int_\tau^T B_j^{-1-\alpha}\Big(1-\alpha\log B_j\Big)\, g(t)\,d(t)\;, \\
I_{11} &= \mathrm{E}\left[\frac{1}{n}\sum_{j=1}^{n_2}\frac{1}{B_j}\right] = \int_\tau^T \frac{1}{B_j}\; g(t)\,d(t)\;, \\
I_{12} &= \mathrm{E}\left[\frac{1}{n}\sum_{j=1}^{n_2}\frac{F_j}{B_j}\right] = \int_\tau^T \frac{F_j}{B_j}\; g(t)\,d(t)\;, \\
I_{13} &= \mathrm{E}\left[\frac{1}{n}\sum_{j=1}^{n_2} F_j\, B_j^{-1-\alpha}\Big(1-\alpha\log B_j\Big)\right] \\
&= \int_\tau^T \sum_{j=1}^{n_2} F_j\, B_j^{-1-\alpha}\Big(1-\alpha\log B_j\Big)\, g(t)\,d(t)\;.
\end{aligned}
$$

APPENDIX II

Detailed calculations of $e_i = \mathrm{E}\left[\frac{n_i}{n}\right]$, $i = 1, 2$ is demonstrated through the following three steps:

At the first step, n new products are tested at stress levels S_1 until time τ, where the test units are assumed independent and identically distributed. The life of items follows the CDF of t in equation (2.2). The number of failures n_1 in time τ is a binomial random variable with parameters n and p_1. From the equation (2.3), we have:

$$\begin{aligned} p_1 &= G(\tau) = \exp\left\{-\left(\frac{\tau}{\theta_1}\right)^{-\alpha}\right\}, \\ e_1 &= \mathrm{E}\left[\frac{n_1}{n}\right] = p_1 = \exp\left\{-\left(\frac{\tau}{\theta_1}\right)^{-\alpha}\right\}. \end{aligned}$$

The second step starts with $n - n_1$ unfailed items, tested at stress levels S_2 until time T. The life of items follows the CDF of t given by the equation (2.2), where the number of failures n_2 follows a binomial distribution with parameters $n - n_1$ and p_2. Then, from the equation (2.2), we have:

$$\begin{aligned} p_2 &= P_r\Big(\textit{item fails in time } T \;\Big|\; \textit{it not fails in time } \tau \textit{ in first step}\Big) \\ &= 1 - P_r\Big(\textit{item not fails in item } T \;\Big|\; \textit{item not fails in time } \tau\Big) \\ &= \frac{\exp\Big\{-\big(\frac{\tau}{\theta_1} + \frac{T-\tau}{\theta_2}\big)^{-\alpha}\Big\} - \exp\Big\{-\big(\frac{\tau}{\theta_1}\big)^{-\alpha}\Big\}}{1 - \exp\Big\{-\big(\frac{\tau}{\theta_1}\big)^{-\alpha}\Big\}}, \\ e_2 &= \mathrm{E}\left[\frac{n_2}{n}\right] = \mathrm{E}\left[\frac{n_2}{n-n_1} \cdot \frac{n-n_1}{n}\right] = p_2 \cdot (1 - p_1). \end{aligned}$$

APPENDIX III

In statistics, the delta method is a result concerning the approximate probability distribution for a function of an asymptotically normal statistical estimator from knowledge of the limiting variance of that estimator.

A consistent estimator B converges in probability to its true value β, and often a central limit theorem can be applied to obtain asymptotic normality:

$$\sqrt{n}(B-\beta) \ \xrightarrow{D} \ N(0,\Sigma) \ ,$$

where n is the number of observations and $\sum$ is a (symmetric positive semi-definite) covariance matrix. Suppose we want to estimate the variance of a function h of the estimator B. Keeping only the first two terms of the Taylor series, and using vector notation for the gradient, we can estimate $h(B)$ as

$$h(B) \ \approx \ h(\beta) + \nabla h(\beta)^T (B-\beta) \ ,$$

which implies the variance of $h(B)$ is approximately

$$\begin{aligned}
\operatorname{Var}\big(h(B)\big) \ &\approx \ \operatorname{Var}\Big(h(\beta) + \nabla h(\beta)^T (B-\beta)\Big) \\
&= \ \operatorname{Var}\Big(h(\beta) + \nabla h(\beta)^T B - \nabla h(\beta)^T \beta\Big) \\
&= \ \operatorname{Var}\Big(\nabla h(\beta)^T B\Big) \\
&= \ \nabla h(\beta)^T \ \operatorname{cov}(\beta) \ \nabla h(\beta) \ .
\end{aligned}$$

One can use the mean value theorem (for real-valued functions of many variables) to see that this does not rely on taking first order approximation.

The delta method therefore implies that

$$\sqrt{n}\Big(h(B) - h(\beta)\Big) \ \xrightarrow{D} \ N\Big(0, \nabla h(\beta)^T \, \Sigma \ \nabla h(\beta)\Big) \, .$$

REFERENCES

[1] Alhadeed, A. A. and Yang, S. S. (2005). Optimal simple step-stress plan for cumulative exposure model using log-normal distribution, *IEEE Transactions on Reliability*, **54**, 64–68.

[2] Alhadeed, A. A. and Yang, S. S. (2002). Optimal simple step-stress plan for Khamis–Higgins Model, *IEEE Trans. Reliability*, **51**(2), 212–215.

[3] Arul, A. J.; Kumar, C. S.; Marimuthu, S. and Singh, O. P. (2003). The power law character of off-site power failures, *Annals of Nuclear Energy*, **30**, 1401–1408.

[4] Bai, D. S.; Kim, M. S. and Lee, S. H. (1989). Optimum simple step-stress accelerated life tests with censoring, *IEEE Transactions on Reliability*, **38**, 528–532.

[5] Barreto-Souza, W.; Cordeiro, G. M. and Simas, A. B. (2011). Some results for beta Fréchet distribution, *Communications in Statistics—Theory and Methods*, **40**, 798–811.

[6] Khamis, I. H. (1997). Optimum M-step design with K stress variables, *Communications in Statistics: Simulation & Computation*, **26**(4), 1301–1313.

[7] Khamis, I. H. and Higgins, J. J. (1996). An alternative to the Weibull cumulative exposure model, *1996 Proc. Section on Quality & Productivity, American Statistical Association, Chicago, Aug.*, **23**(4), 230–350.

[8] Kotz, S. and Nadarajah, S. (2000). *Extreme Value Distributions: Theory and Applications*, London: Imperial college press.

[9] Fard, N. and Li, C. (2009). Optimal simple step stress accelerated life test design for reliability prediction, *Journal of Statistical Planning and Inference*, **139**, 1799–1808.

[10] Miller, R. and Nelson, W. (1983). Optimum simple step-stress plans for accelerated life testing, *IEEE Transactions on Reliability*, **32**(1), 59–65.

[11] Krishna, E.; Jose, K. K. and Ristic, M. M. (2013). Applications of Marshall-Olkin Fréchet distribution, *Communications in Statistics—Simulation and Computation*, **42**, 76-89.

[12] Meeker, W. Q. and Escobar, L. A. (1998). *Statistical methods for reliability data*, A Wiley-Interscience publication.

[13] Miller, R. and Nelson, W. B. (1983). Optimum simple step-stress plans for accelerated life testing, *IEEE Transactions on Reliability*, **32**, 59-65.

[14] Nelson, W. B. (1980). Accelerated life testing-step-stress models and data analyses, *IEEE Transactions on Reliability*, **29**, 103-108

[15] Nelson, W. B. (1990). *Accelerated Life Testing: Statistical Models, Data Analysis and Test Plans*, John Wiley and sons, New York.

[16] Srivastava, P. W. and Shukla, R. (2008). A log-logistic step-stress model, *IEEE Transactions on Reliability*, **57**, 431-434.

REVSTAT – Statistical Journal
Volume 15, Number 1, January 2017, 25–43

EXPANSIONS FOR QUANTILES AND MULTIVARIATE MOMENTS OF EXTREMES FOR HEAVY TAILED DISTRIBUTIONS

Authors: Christopher Withers
– Industrial Research Limited,
Lower Hutt, New Zealand

Saralees Nadarajah
– School of Mathematics, University of Manchester,
Manchester M13 9PL, UK
mbbsssn2@manchester.ac.uk

Received: August 2014 Revised: March 2015 Accepted: May 2015

Abstract:

- Let $X_{n,r}$ be the r-th largest of a random sample of size n from a distribution function $F(x) = 1 - \sum_{i=0}^{\infty} c_i\, x^{-\alpha-i\beta}$ for $\alpha > 0$ and $\beta > 0$. An inversion theorem is proved and used to derive an expansion for the quantile $F^{-1}(u)$ and powers of it. From this an expansion in powers of $\left(n^{-1}, n^{-\beta/\alpha}\right)$ is given for the multivariate moments of the extremes $\{X_{n,n-s_i},\, 1 \le i \le k\}/n^{1/\alpha}$ for fixed $\mathbf{s} = (s_1, ..., s_k)$, where $k \ge 1$. Examples include the Cauchy, Student's t, F, second extreme distributions and stable laws of index $\alpha < 1$.

Key-Words:

- *Bell polynomials; extremes; inversion theorem; moments; quantiles.*

AMS Subject Classification:

- 62E15, 62E17.

1. INTRODUCTION

For $1 \le r \le n$, let $X_{n,r}$ be the r-th largest of a random sample of size n from a continuous distribution function F on $\mathbb{R}$, the real numbers. Let f denote the density function of F when it exists.

The study of the asymptotes of the moments of $X_{n,r}$ has been of considerable interest. McCord [12] gave a first approximation to the moments of $X_{n,1}$ for three classes. This showed that a moment of $X_{n,1}$ can behave like any positive power of n or $n_1 = \log n$. (Here, log is to the base e.) Pickands [15] explored the conditions under which various moments of $\left(X_{n,1} - b_n\right)/a_n$ converge to the corresponding moments of the extreme value distribution. It was proved that this is indeed true for all F in the domain of attraction of an extreme value distribution provided that the moments are finite for sufficiently large n. Nair [13] investigated the limiting behavior of the distribution and the moments of $X_{n,1}$ for large n when F is the standard normal distribution function. The results provided rates of convergence of the distribution and the moments of $X_{n,1}$. Downey [4] derived explicit bounds for $\mathbb{E}\left[X_{n,1}\right]$ in terms of the moments associated with F. The bounds were given up to the order $o\left(n^{1/\rho}\right)$, where $\displaystyle\int_{-\infty}^{\infty} |x|^\rho \, dF(x)$ is defined, so $\mathbb{E}\left[X_{n,1}\right]$ grows slowly with the sample size. For other work, we refer the readers to Ramachandran [16], Hill and Spruill [9] and Hüsler *et al.* [10].

The main aim of this paper is to study multivariate moments of $\{X_{n,n-s_i},\ 1 \le i \le k\}$ for fixed $\mathbf{s} = (s_1, ..., s_k)$, where $k \ge 1$. We suppose F is heavy tailed, *i.e.*,

$$1 - F(x) \sim C x^{-\alpha} \tag{1.1}$$

as $x \to \infty$ for some $C > 0$ and $\alpha > 0$. For a nonparametric estimate of α, see Novak and Utev [14].

There are many practical examples giving rise to $\{X_{n,n-s_i},\ 1 \le i \le k\}$ for heavy tailed F. Perhaps the most prominent example is the Hill's estimator (Hill [8]) for the *extremal index* given by

$$-\log X_{n,n-k} + k^{-1} \sum_{i=1}^{k} \log X_{n,n-i+1} \,.$$

Clearly, this is a function of $X_{n,n-s_i}$, $1 \le i \le k$. Real life applications of the Hill's estimator are far too many to list.

Since Hill [8], many other estimators have been proposed for the extremal index, see Gomes and Guillou [6] for an excellent review of such estimators. Each of these estimators is a function of $X_{n,n-s_i}$, $1 \le i \le k$. No doubt that many more estimators taking the form of a function of $X_{n,n-s_i}$, $1 \le i \le k$ will be proposed in the future.

A possible application of the results in this paper is to assess optimality of these estimators. Suppose we can write the general form of the estimators as

$$\omega \;=\; \omega\big(X_{n,n-s_1}, X_{n,n-s_2}, ..., X_{n,n-s_k}; \boldsymbol{\mu}\big) \;, \tag{1.2}$$

where $\boldsymbol{\mu}$ contains some parameters, which include k itself. The optimum values of $\boldsymbol{\mu}$ can be based on criteria like bias and mean squared error. For example, $\boldsymbol{\mu}$ could be chosen as the value minimizing the bias of ω or the value minimizing the mean squared error of ω. If (1.2) can be expanded as

$$\omega \;=\; \sum_{\theta_1,\theta_2,...,\theta_k} a\big(\theta_1,\theta_2,...,\theta_k;\boldsymbol{\mu}\big) \prod_{i=1}^{k} X_{n,n-s_i}^{\theta_i}$$

then the bias and mean squared error of ω can be expressed as

$$\mathrm{Bias}(\omega) \;=\; \sum_{\theta_1,\theta_2,...,\theta_k} a\big(\theta_1,\theta_2,...,\theta_k;\boldsymbol{\mu}\big)\; \mathbb{E}\left[\prod_{i=1}^{k} X_{n,n-s_i}^{\theta_i}\right] - \,\omega$$

and

$$\begin{aligned}\mathrm{MSE}(\omega) \;=\; &\sum_{\theta_1,\theta_2,...,\theta_k}\;\sum_{\vartheta_1,\vartheta_2,...,\vartheta_k} a\big(\theta_1,\theta_2,...,\theta_k;\boldsymbol{\mu}\big)\, a\big(\vartheta_1,\vartheta_2,...,\vartheta_k;\boldsymbol{\mu}\big)\; \mathbb{E}\left[\prod_{i=1}^{k} X_{n,n-s_i}^{\theta_i+\vartheta_i}\right] \\ &- \left\{\sum_{\theta_1,\theta_2,...,\theta_k} a\big(\theta_1,\theta_2,...,\theta_k;\boldsymbol{\mu}\big)\; \mathbb{E}\left[\prod_{i=1}^{k} X_{n,n-s_i}^{\theta_i}\right]\right\}^2 + \big[\mathrm{Bias}(\omega)\big]^2 \,,\end{aligned}$$

respectively. Both involve multivariate moments of $X_{n,n-s_i}$, $1 \le i \le k$. Expressions for the latter are given in Section 2, in particular, Theorem 2.2. Hence, general estimators can be developed for $\boldsymbol{\mu}$ which minimize bias, mean squared error, etc. Such developments could apply to any future estimator (also to any past estimator) of the extremal index taking the form of (1.2).

Note that $U_{n,r} = F\,(X_{n,r})$ is the r-th order statistics from $U(0,1)$. For $1 \le r_1 < r_2 < \cdots < r_k \le n$ set $U_{n,\mathbf{r}} = \{U_{n,r_i}, 1 \le i \le k\}$. By Section 14.2 of Stuart and Ord [17], $U_{n,\mathbf{r}}$ has the multivariate beta density function

$$U_{n,\mathbf{r}} \sim B\,(\mathbf{u}:\mathbf{r}) \;=\; \prod_{i=0}^{k} (u_{i+1}-u_i)^{r_{i+1}-r_i-1} \big/ B_n(\mathbf{r}) \tag{1.3}$$

on $0 < u_1 < \cdots < u_k < 1$, where $u_0 = 0$, $u_{k+1} = 1$, $r_0 = 0$, $r_{k+1} = n+1$ and

$$B_n(\mathbf{r}) \;=\; \prod_{i=1}^{k} B\,(r_i, r_{i+1}-r_i) \;. \tag{1.4}$$

David and Johnson [3] expanded $X_{n,r_i} = F^{-1}(U_{n,r_i})$ about $u_{n,i} = \mathbb{E}[U_{n,r_i}] = r_i/(n+1)$: $X_{n,r_i} = \sum_{j=0}^{\infty} G^{(j)}(u_{n,i})\,(U_{n,i} - u_{n,i})^j / j!$, where $G(u) = F^{-1}(u)$ and $G^{(j)}(u) = d^j G(u)/du^j$, and using the properties of (1.3) showed that if $\mathbf{r}$ depends on n in such a ways that $\mathbf{r}/n \to \mathbf{p} \in (\mathbf{0}, \mathbf{1})$ as $n \to \infty$ then the m-th order cumulants of $X_{n,\mathbf{r}} = \{X_{n,r_i}, 1 \le i \le k\}$ have magnitude $O\left(n^{1-m}\right)$ — at least for $n \le 4$, so that the distribution function of $X_{n,\mathbf{r}}$ has a multivariate Edgeworth expansion in powers of $n^{-1/2}$. (Alternatively one can use James and Mayne [11] to derive the cumulants of $X_{n,\mathbf{r}}$ from those of $U_{n,\mathbf{r}}$.) The method requires the derivatives of F at $\left\{F^{-1}(p_i), 1 \le i \le k\right\}$ so breaks down if $p_i = 0$ or $p_k = 1$ — which is the situation we study here.

In Withers and Nadarajah [18], we showed that for fixed $\mathbf{r}$ when (1.1) holds the distribution of $X_{n,n\mathbf{1}-\mathbf{r}}$ (where $\mathbf{1}$ is the vector of ones in $\mathbb{R}^k$), suitably normalized tends to a certain multivariate extreme value distribution as $n \to \infty$, and so obtained the leading terms of the expansions of its moments in inverse powers of n. Here, we show how to extend those expansions when

$$F^{-1}(u) = \sum_{i=0}^{\infty} b_i (1-u)^{\alpha_i} \tag{1.5}$$

with $\alpha_0 < \alpha_1 < \cdots$, that is, $\{1 - F(x)\}\, x^{-1/\alpha_0}$ has a power series in $\{x^{-\delta_i} \colon \delta_i = (\alpha_i - \alpha_0)/\alpha_0\}$. Hall [7] considered (1.5) with $\alpha_i = i - 1/\alpha$, but did not give the corresponding expansion for $F(x)$ or expansions in inverse powers of n. He applied it to the Cauchy. In Section 2, we demonstrate the method when

$$1 - F(x) = x^{-\alpha} \sum_{i=0}^{\infty} c_i\, x^{-i\beta}\,, \tag{1.6}$$

where $\alpha > 0$ and $\beta > 0$. In this case, (1.5) holds with $\alpha_i = (i\beta - 1)/\alpha$. In Section 3, we apply it to the Student's t, F and second extreme value distributions and to stable laws of exponent $\alpha < 1$. The appendix gives the inverse theorem needed to pass from (1.6) to (1.5), and expansions for powers and logs of series.

We use the following notation and terminology. Let $(x)_i = \Gamma(x+i)/\Gamma(x)$ and $\langle x\rangle_i = \Gamma(x+1)/\Gamma(x-i+1)$. An inequality in $\mathbb{R}^k$ consists of k inequalities. For example, for $\mathbf{x}$ in $\mathbb{C}^k$, where $\mathbb{C}$ is the set of complex numbers, $\mathrm{Re}(\mathbf{x}) < \mathbf{0}$ means that $\mathrm{Re}(x_i) < 0$ for $1 \le i \le k$. Also let $I(A) = 1$ if A is true and $I(A) = 0$ if A is false. For $\boldsymbol{\theta} \in \mathbb{C}^k$ let $\overline{\boldsymbol{\theta}}$ denote the vector with $\overline{\theta}_i = \sum_{j=1}^{i} \theta_j$.

2. MAIN RESULTS

For $1 \le r_1 < \cdots < r_k \le n$ set $s_i = n - r_i$. Here, we show how to obtain expansions in inverse powers of n for the moments of the $X_{n,\mathbf{s}}$ for fixed $\mathbf{r}$ when (1.5) holds, and in particular when the upper tail of F satisfies (1.6).

Theorem 2.1. *Suppose (1.6) holds with c_0, α, $\beta > 0$. Then $F^{-1}(u)$ is given by (1.5) with $\alpha_i = ia - 1/\alpha$, $a = \beta/\alpha$ and $b_i = C_{i,1/\alpha}$, where $C_{i,\psi} = c_0^{\psi}\widehat{C}_i(-\psi, c_0, \mathbf{x}^*)$ of (A.3) and $x_i^* = x_i^*(a, 1, \mathbf{c})$ of (A.4). In particular,*

$$\begin{aligned}
C_{0,\psi} &= c_0^{\psi} ,\\
C_{1,\psi} &= \psi\, c_0^{\psi-a-1} c_1 ,\\
C_{2,\psi} &= \psi\, c_0^{\psi-2a-2}\left\{c_0 c_2 + (\psi - 2a - 1)\, c_1^2/2\right\} ,\\
C_{3,\psi} &= \psi\, c_0^{\psi-3a-3}\Big[c_0^2 c_2 + (\psi-3a-1)\, c_0 c_1 c_2 + \left\{(\psi+1)_2/6(\psi+3a/2)(a+1)\right\} c_1^3\Big] ,
\end{aligned}$$

and so on. Also for any θ in $\mathbb{R}$,

$$\left\{F^{-1}(u)\right\}^{\theta} = \sum_{i=0}^{\infty}(1-u)^{ia-\psi} C_{i,\psi} \tag{2.1}$$

at $\psi = \theta/\alpha$.

On those rate occasions, where the coefficients $d_i = C_{i,1/\alpha}$ in $F^{-1}(u) = \sum_{i=0}^{\infty}(1-u)^{ia-1/\alpha} d_i$ are known from some alternative formula then one can use $C_{i,\psi} = d_0^{\theta}\widehat{C}_i(\theta, 1/d_0, \mathbf{d})$ of (A.3).

Proof of Theorem 2.1: By Theorem A.1 with $k = 1$, we have $x^{-\alpha} = \sum_{i=0}^{\infty} x_i^*(1-u)^{1+ia}$ at $u = F(x)$, where

$$\begin{aligned}
x_0^* &= c_0^{-1} ,\\
x_1^* &= c_0^{-a-2} c_1 ,\\
x_2^* &= c_0^{-2a-3}\left\{-c_0 c_2 + (a+1) c_1^2\right\} ,\\
x_3^* &= c_0^{-3a-4}\left\{-c_0^2 c_3 + (2+3a)\, c_0 c_1 c_2 - (2+3a)\,(1+a)\, c_1^2/2\right\} ,
\end{aligned}$$

and so on. So, for S of (A.1), $x^{-\alpha} = c_0^{-1} v\,[1 + c_0 S(v^a, \mathbf{x}^*)]$ at $v = 1-u$. Now apply (A.2). $\square$

Lemma 2.1. *For $\boldsymbol{\theta}$ in $\mathbb{C}^k$,*

$$\mathbb{E}\left[\prod_{i=1}^{k}(1-U_{n,r_i})^{\theta_i}\right] = b_n\left(\mathbf{r}:\overline{\boldsymbol{\theta}}\right) , \tag{2.2}$$

where

$$b_n\left(\mathbf{r}:\overline{\boldsymbol{\theta}}\right) = \prod_{i=1}^{k} b\left(r_i-r_{i-1}, n-r_i+1:\overline{\theta}_i\right) \tag{2.3}$$

and $b(\alpha,\beta:\theta) = B(\alpha,\beta+\theta)/B(\alpha,\beta)$. Also in (1.4),

$$B_n(\mathbf{r}) = \prod_{i=1}^{k} B(r_i-r_{i-1}, n-r_i+1) . \tag{2.4}$$

Since $B(\alpha,\beta)=\infty$ for $\mathrm{Re}\beta \le 0$, for (2.2) to be finite we need $n-r_i+1+\mathrm{Re}\,\overline{\theta}_i > 0$ for $1 \le i \le k$.

Proof of Lemma 2.1: Let I_k denote the left hand side of (2.2). Then $I_k = \displaystyle\int B_n(\mathbf{u}:\mathbf{r})\prod_{i=1}^{k}(1-u_i)^{\theta_i}\,du_1\cdots du_k$ integrated over $0<u_1<\cdots<u_k<1$ by (1.3). So, (2.2), (2.4) hold for $k=1$. Set $s_i = (u_i-u_{i-1})\,/\,(1-u_{i-1})$. Then

$$I_2 = \int_0^1 u_1^{r_1-1}(1-u_1)^{\theta_1}\int_{u_1}^1 (u_2-u_1)^{r_2-r_1-1}(1-u_2)^{r_3-r_2-1+\theta_2}\,du_2\big/B_n(\mathbf{r}) ,$$

which is the the right hand side of (2.2) with denominator replaced by the right hand side of (2.3). Putting $\boldsymbol{\theta}=\mathbf{0}$ gives (2.2), (2.4) for $k=2$. Now use induction. □

Lemma 2.2. *In Lemma 2.1, the restriction*

$$1 \le r_1 < \cdots < r_k \le n \quad \textit{may be relaxed to} \quad 1 \le r_1 \le \cdots \le r_k \le n . \tag{2.5}$$

Proof: For $k=2$, the second factor in the right hand side of (2.3) is $b(r_2-r_1,\ n-r_2+1:\overline{\theta}_2) = f\left(\overline{\theta}_2\right)/f(0)$, where $f\left(\overline{\theta}_2\right) = \Gamma\left(n-r_2+1+\overline{\theta}_2\right)/\Gamma\left(n-r_1+1+\overline{\theta}_2\right) = 1$ if $r_2=r_1$ and the first factor is $b\left(r_1, n-r_1+1:\overline{\theta}_1\right) = \mathbb{E}\left[(1-U_{n,r_1})^{\overline{\theta}_1}\right]$. Similarly, if $r_i=r_{i-1}$, the i-th factor is 1 and the product of the others is $\mathbb{E}\left[\displaystyle\prod_{j=1,j\neq i}^{k}\left(1-U_{n,r_j}\right)^{\theta_j^*}\right]$, where $\theta_j^*=\theta_j$ for $j\neq i-1$ and $\theta_j^*=\theta_{i-1}+\theta_i$ for $j=i-1$. □

Corollary 2.1. *In any formulas for $\mathbb{E}\left[g\left(X_{n,\mathbf{r}}\right)\right]$ for some function g, (2.5) holds. In particular it holds for the moments and cumulants of $X_{n,\mathbf{r}}$.*

This result is very important as it means we can dispense with treating the 2^{k-1} cases $r_i < r_{i+1}$ or $r_i = r_{i+1}$, $1 \le i \le k-1$ separately. For example, Hall [7] treats the two cases for $\cos\left(X_{n,\mathbf{r}}, X_{n,\mathbf{s}}\right)$ separately and David and Johnson [3] treat the 2^{k-1} cases for the k-th order cumulants of $X_{n,\mathbf{r}}$ separately for $k \le 4$.

Theorem 2.2. *Under the conditions of Theorem 2.1,*

$$\mathbb{E}\left[\prod_{i=1}^{k} X_{n,r_i}^{\theta_i}\right] = \sum_{i_1,\ldots,i_k=0}^{\infty} C_{i_1,\psi_1} \cdots C_{i_k,\psi_k} b_n\left(\mathbf{r} : \bar{\mathbf{i}} a - \overline{\boldsymbol{\theta}}/\alpha\right) \tag{2.6}$$

with b_n as in (2.3), where $\boldsymbol{\psi} = \boldsymbol{\theta}/\alpha$. All terms are finite if $\operatorname{Re}\overline{\boldsymbol{\theta}} < (\mathbf{s}+1)\,\alpha$, where $s_i = n - r_i$.

Lemma 2.3. *For α, β positive integers and θ in $\mathbb{C}$,*

$$b(\alpha, \beta : \theta) = \prod_{j=\beta}^{\alpha+\beta-1} (1 + \theta/j)^{-1}\ . \tag{2.7}$$

So, for $\boldsymbol{\theta}$ in $\mathbb{C}^k$,

$$b_n\left(\mathbf{r} : \overline{\boldsymbol{\theta}}\right) = \prod_{i=1}^{k} \prod_{j=s_i+1}^{s_{i-1}} \left(1 + \overline{\theta}_i/j\right)^{-1}\ , \tag{2.8}$$

where $s_i = n - r_i$ and $r_0 = 0$.

Proof: The left hand side of (2.7) is equal to $\Gamma(\beta+\theta)\Gamma(\alpha+\beta)\,/\,\{\Gamma(\beta+\theta+\alpha)\Gamma(\beta)\}$. But $\Gamma(\alpha+x)/\Gamma(x) = (x)_\alpha$, so (2.7) holds, and hence (2.8). $\square$

From (2.3) we have, interpreting $\prod_{i=2}^{k-1} b_i$ as 1,

Lemma 2.4. *For $s_i = n - r_i$,*

$$b_n\left(\mathbf{r} : \overline{\boldsymbol{\theta}}\right) = B\left(\mathbf{s} : \overline{\boldsymbol{\theta}}\right) n!\,/\,\Gamma\left(n + 1 + \overline{\theta}_1\right), \tag{2.9}$$

where

$$B\left(\mathbf{s} : \overline{\boldsymbol{\theta}}\right) = \Gamma\left(s_1 + 1 + \overline{\theta}_1\right)(s_1!)^{-1} \prod_{i=2}^{k} b\left(s_{i-1} - s_i, s_i + 1 : \overline{\theta}_i\right)$$

does not depend on n for fixed $\mathbf{s}$.

Lemma 2.5. *We have*

$$n!\,/\,\Gamma(n+1+\theta) \;=\; n^{-\theta}\sum_{i=0}^{\infty} e_i(\theta) n^{-i}\;,$$

where

$$\begin{aligned}
&e_0(\theta) = 1\;, \quad e_1(\theta) = -(\theta)_2/2\;, \quad e_2(\theta) = (\theta)_3\,(3\theta+1)/24\;,\\
&e_3(\theta) = -(\theta)_4\,(\theta)_2/\,(4!\cdot 2)\;, \quad e_4(\theta) = (\theta)_5\,\big(15\theta^3+30\theta^2+5\theta-2\big)\,/\,(5!\cdot 48)\;,\\
&e_5(\theta) = -(\theta)_6\,(\theta)_2\,\big(3\theta^2+7\theta-2\big)\,/\,(6!\cdot 16)\;,\\
&e_6(\theta) = (\theta)_7\,\big(63\theta^5+315\theta^4+315\theta^3-91\theta^2-42\theta+16\big)\,/\,(7!\cdot 576)\;,\\
&e_7(\theta) = -(\theta)_8\,(\theta)_2\,\big(9\theta^4+54\theta^3+51\theta^2-58\theta+16\big)\,/\,(8!\cdot 144)\;.
\end{aligned}$$

Proof: Apply equation (6.1.47) of Abramowitz and Stegun [1]. □

So, (2.6), (2.9) yield the joint moments of $X_{n,\mathbf{r}}n^{-1/\alpha}$ for fixed $\mathbf{s}$ as a power series in $(1/n, n^{-\alpha})$:

Corollary 2.2. *Under the conditions of Theorem 2.1,*

$$\mathbb{E}\left[\prod_{i=1}^{k} X_{n,n-s_i}^{\theta_i}\right] = \sum_{j=0}^{\infty} n!\,\Gamma\big(n+1+ja-\overline{\psi}_1\big)^{-1}\,C_j\,(\mathbf{s}:\boldsymbol{\psi})\;, \tag{2.10}$$

where $\boldsymbol{\psi} = \boldsymbol{\theta}/\alpha$ *and*

$$C_j\,(\mathbf{s}:\boldsymbol{\psi}) \;=\; \sum\Big\{C_{i_1,\psi_1}\cdots C_{i_k,\psi_k} B\,\big(\mathbf{s}:\bar{\mathbf{i}}a-\overline{\boldsymbol{\psi}}\big):\; i_1+\cdots+i_k=j\Big\}\;.$$

So, if $\mathbf{s}$, $\boldsymbol{\theta}$ are fixed as $n\to\infty$ and $\operatorname{Re}\big(\overline{\boldsymbol{\theta}}\big) < (\mathbf{s}+\mathbf{1})\,\alpha$, then the left hand side of (2.10) is equal to

$$n^{\psi_1}\sum_{i,j=0}^{\infty} n^{-i-ja}\,e_i\big(ja-\overline{\psi}_1\big)\,C_j\,(\mathbf{s}:\boldsymbol{\psi})\;. \tag{2.11}$$

If a is rational, say $a = M/N$ then the left hand side of (2.10) is equal to

$$n^{\overline{\psi}_1}\sum_{m=0}^{\infty} n^{-m/N} d_m\,(\mathbf{s}:\boldsymbol{\psi})\;, \tag{2.12}$$

where

$$\begin{aligned}
d_m\,(\mathbf{s}:\boldsymbol{\psi}) \;&=\; \sum\Big\{e_i\,\big(ja-\overline{\psi}_1\big)\,C_j\,(\mathbf{s}:\boldsymbol{\psi}):\; iN+jM=m\Big\}\\
&=\; \sum\Big\{e_{m-ja}\,\big(ja-\overline{\psi}_1\big)\,C_j\,(\mathbf{s}:\boldsymbol{\psi}):\; 0\le j\le m/a\Big\}
\end{aligned}$$

if $N=1$; so for d_m to depend on c_1 and not just c_0 we need $m\le M$.

The leading term in (2.11) does not involve c_1 so may be deduced from the multivariate extreme value distribution that the law of $X_{n,n-s_i}$, suitably normalized, tends to. The same is true of the leading terms of its cumulants. See Withers and Nadarajah [18] for details.

The leading terms in (2.11) are

$$n^{\overline{\psi}_1}\left[\left\{1-n^{-1}\langle\overline{\psi}_1\rangle_2/2\right\}C_0\left(\mathbf{s}:\boldsymbol{\psi}\right)+n^{-a}C_1\left(\mathbf{s}:\boldsymbol{\psi}\right)+O\left(n^{-2a_0}\right)\right],$$

where

$$\begin{aligned} a_0 &= \min(a,1)\,, \\ C_0\left(\mathbf{s}:\boldsymbol{\psi}\right) &= c_0B\left(\mathbf{s}:-\overline{\boldsymbol{\psi}}\right)\,, \\ C_1\left(\mathbf{s}:\boldsymbol{\psi}\right) &= c_0^{\overline{\psi}_1-a-2}c_1\sum_{j=1}^{k}\psi_jB\left(\mathbf{s}:a\mathbf{I}_j-\overline{\boldsymbol{\psi}}\right) \end{aligned}$$

and $I_{j,m}=I(m\leq j)$. For $k=1$,

$$\begin{aligned} C_0(s:\psi) &= c_0^{\psi}(s+1)_{-\psi} = c_0^{\psi}/\langle s\rangle_{\psi}\,, \\ C_1(s:\psi) &= \psi c_0^{\psi-a-1}c_1(s+1)_{a-\psi} = \psi c_0^{\psi-a-1}c_1/\langle s\rangle_{\psi-a}\,. \end{aligned}$$

Set

$$\pi_{\mathbf{s}}(\lambda) = b\left(s_1-s_2,s_2+1:\lambda\right) = \prod_{j=s_2+1}^{s_1}1/\left(1+\lambda/j\right)$$

for λ an integer. For example, $\pi_{\mathbf{s}}(1)=(s_2+1)/(s_1+1)$ and $\pi_{\mathbf{s}}(-1)=s_1/s_2$. Then for $k=2$,

$$\begin{aligned} C_0\left(\mathbf{s}:\lambda\mathbf{1}\right) &= c_0^{2\lambda}\langle s_1\rangle_{2\lambda}^{-1}\,\pi_{\mathbf{s}}(-\lambda) \\ &= c_0^{2}\left(s_1-1\right)^{-1}s_2 \quad \text{for } \lambda=1 \\ &= c_0^{2}\langle s_2-2\rangle_2^{-1}\langle s_2\rangle_2^{-1} \quad \text{for } \lambda=2 \end{aligned}$$

and

$$\begin{aligned} C_1\left(\mathbf{s}:\lambda\mathbf{1}\right) &= \lambda\,c_0^{2\lambda-a-1}c_1\langle s_1\rangle_{2\lambda-a}^{-1}\left\{\pi_{\mathbf{s}}(-\lambda)+\pi_{\mathbf{s}}(a-\lambda)\right\} \\ &= \lambda\,c_0^{1-a}c_1\langle s_1\rangle_{2-a}^{-1}\left\{s_1/s_2+\pi_{\mathbf{s}}(a-1)\right\} \quad \text{for } \lambda=1 \\ &= \lambda\,c_0^{3-a}c_1\langle s_1\rangle_{4-a}^{-1}\left\{\langle s_1\rangle_2\langle s_2\rangle_2^{-1}+\pi_{\mathbf{s}}(a-2)\right\} \quad \text{for } \lambda=2\,. \end{aligned}$$

Set $\lambda=1/\alpha$, $Y_{n,s}=X_{n,n-s}/\left(nc_0\right)^{\lambda}$ and $E_{\mathbf{c}}=\lambda\,c_0^{-a-1}c_1$. Then for $s>\lambda-1$

$$\mathbb{E}\left[Y_{n,s}\right] = \left\{1-n^{-1}\langle\lambda\rangle_2/2\right\}\langle s\rangle_{\lambda}^{-1}+n^{-a}E_{\mathbf{c}}\langle s\rangle_{\lambda-a}^{-1}+O\left(n^{-2a_0}\right) \tag{2.13}$$

and for $s_1>2\lambda-1$, $s_2>\lambda-1$, $s_1\geq s_2$,

$$\mathbb{E}\left[Y_{n,s_1}Y_{n,s_2}\right] = \left\{1-n^{-1}\langle 2\lambda\rangle_2/2\right\}B_{2,0}+n^{-a}E_{\mathbf{c}}D_a+O\left(n^{-2a_0}\right)\,, \tag{2.14}$$

where $B_{2,0} = \langle s_1 \rangle_{2\lambda}^{-1} \pi_{\mathbf{s}}(-\lambda)$, $D_a = \langle s_1 \rangle_{2\lambda-a}^{-1} \{\pi_{\mathbf{s}}(-\lambda) + \pi_{\mathbf{s}}(a-\lambda)\}$ and

$$\text{(2.15)} \qquad \mathrm{Cov}\,(Y_{n,s_1}, Y_{n,s_2}) \;=\; F_0 + F_1/n + E_{\mathbf{c}} F_2/n + O\left(n^{-2a_0}\right) ,$$

where $F_0 = B_{2,0} - \langle s_1 \rangle_\lambda^{-1} \langle s_2 \rangle_\lambda^{-1}$, $F_1 = \langle \lambda \rangle_2 \langle s_1 \rangle_\lambda^{-1} \langle s_2 \rangle_\lambda^{-1} - \langle 2\lambda \rangle_2 B_{2,0}/2$ and $F_2 = D_a - \langle s_1 \rangle_\lambda^{-1} \langle s_2 \rangle_{\lambda-a}^{-1} - \langle s_1 \rangle_{\lambda-a}^{-1} \langle s_2 \rangle_\lambda^{-1}$. Similarly, we may use (2.11) to approximate higher order cumulants. If $a = 1$ this gives $\mathbb{E}\,[Y_{n,\mathbf{s}}]$ and $\mathrm{Cov}\,(Y_{n,s_1}, Y_{n,s_2})$ to $O\left(n^{-2}\right)$.

Example 2.1. Suppose $\alpha = 1$. Then $Y_{n,s} = X_{n,n-s}/\,(nc_0)$, $E_{\mathbf{c}} = c_0^{-a-1} c_1$, $B_{2,0} = -F_1 = (s_1-1)^{-1}\, s_2^{-1}$, $F_0 = \langle s_1 \rangle_2^{-1} s_2^{-1}$, $D_a = \langle s_1 \rangle_{2-a}^{-1} G_a$, where $G_a = s_1 s_2^{-1} + \pi_{\mathbf{s}}(a-1)$ for $s_1 \geq s_2$, $G_a = 2$ for $s_1 = s_2$ and $F_2 = D_a - s_1^{-1} \langle s_2 \rangle_{1-a}^{-1} - s_2^{-1} \langle s_1 \rangle_{1-a}^{-1}$. So,

$$\text{(2.16)} \qquad \mathbb{E}\,[Y_{n,s}] \;=\; s^{-1} + n^{-a} E_{\mathbf{c}} \langle s \rangle_{1-a}^{-1} + O\left(n^{-2a_0}\right)$$

for $s > 0$ and (2.14)–(2.15) hold if

$$\text{(2.17)} \qquad s_1 > 1\,, \quad s_2 > 0\,, \quad s_1 \geq s_2\,.$$

A little calculation shows that $C_0\,(\mathbf{s} : \mathbf{1}) = c_0^k B_{k,0}$, $C_1\,(\mathbf{s} : \mathbf{1}) = c_0^{k-a-1} c_1 B_{k,\cdot}$, and

$$\begin{aligned} \mathbb{E}\left[\prod_{i=1}^k Y_{n,s_i}\right] \;&=\; \left\{1 + n^{-1} \langle k \rangle_2/2\right\} B_{k,0} + n^{-a} E_{\mathbf{c}} B_{k,\cdot} + O\left(n^{-2a_0}\right) \\ &=\; m_0(\mathbf{s}) + n^{-1} m_1(\mathbf{s}) + n^{-a} m_a(\mathbf{s}) + O\left(n^{-2a_0}\right) \end{aligned}$$

say for $s_i > k - i$, $1 \leq i \leq k$ and $s_1 \geq \cdots \geq s_k$, where

$$\begin{aligned} B_{k,\cdot} \;&=\; \sum_{j=1}^k B_{k,j}\,, \\ B_{k,0} \;&=\; \prod_{i=1}^k 1/\,(s_1 - k + i)\,, \\ B_{k,j} \;&=\; \prod_{i=1}^{j-1} (s_i - k + a + i)^{-1} \langle s_j - k + j + 1 \rangle_{a-1} \prod_{i=j+1}^{k} (s_i - k + i)^{-1}\,, \\ B_{k,k} \;&=\; \prod_{i=1}^{k-1} (s_i - k + a + i)^{-1} \langle s_k \rangle_{1-a}^{-1} \end{aligned}$$

for $s_i > k - i$ and $1 \leq j < k$. For example, $B_{1,0} = s_1$, $B_{2m,0} = (s_1 - 1)^{-1}\, s_2^{-1}$ and $B_{3,0} = (s_1 - 2)^{-1} (s_2 - 1)^{-1} s_3^{-1}$. So, $\kappa_n(\mathbf{s}) = \kappa\,(Y_{n,s_1}, ..., Y_{n,s_k})$, the joint cumulant of $(Y_{n,s_1}, ..., Y_{n,s_k})$, is given by $\kappa_n(\mathbf{s}) = \kappa_0(\mathbf{s}) + n^{-1}\kappa_1(\mathbf{s}) + n^{-a}\kappa_a(\mathbf{s}) +$

$O\left(n^{-2a_0}\right)$, where, for example,

$$\begin{aligned}
\kappa_0(s_1,s_2,s_3) &= m_0(s_1,s_2,s_3) - m_0(s_1)\,m_0(s_2,s_3) - m_0(s_2)\,m_0(s_1,s_3)\\
&\quad - m_0(s_3)\,m_0(s_1,s_2) + 2\prod_{i=1}^{3} m_0(s_i)\\
&= 2(s_1+s_2-2)\,D(s_1,s_2,s_3)\ ,
\end{aligned}$$

$$\begin{aligned}
\kappa_1(s_1,s_2,s_3) &= m_1(s_1,s_2,s_3) - m_0(s_1)\,m_1(s_2,s_3) - m_0(s_2)\,m_1(s_1,s_3)\\
&\quad - m_0(s_3)\,m_1(s_1,s_2)\\
&= 2\left\{s_2(1-2s_1)+s_1-s_1^2\right\}/D(s_1,s_2,s_3) \qquad \text{since} \quad m_1(s_1)=0\ ,
\end{aligned}$$

$$\begin{aligned}
\kappa_a(s_1,s_2,s_3) &= m_a(s_1,s_2,s_3) - m_0(s_1)\,m_a(s_2,s_3) - m_a(s_1)\,m_0(s_2,s_3)\\
&\quad - m_0(s_2)\,m_a(s_1,s_3) - m_a(s_2)\,m_0(s_1,s_3) - m_0(s_3)\,m_a(s_1,s_2)\\
&\quad - m_a(s_3)\,m_0(s_1,s_2) + 2\,m_0(s_1)\,m_0(s_2)\,m_a(s_3)\\
&\quad + 2\,m_0(s_3)\,m_0(s_1)\,m_a(s_2) + 2\,m_0(s_2)\,m_0(s_3)\,m_a(s_1)\ ,
\end{aligned}$$

where $D(s_1,s_2,s_3) = \langle s_1\rangle_3\,\langle s_2\rangle_2\,s_3$.

Consider the case $a=1$. Then $\kappa_a(s_1,s_2,s_3)=0$ so

$$\begin{aligned}
\kappa_n(s_1,s_2,s_3) &= 2\Big\{s_1+s_2-2+n^{-1}\left(s_2(1-2s_1)+s_1-s_1^2\right)\Big\}\Big/D(s_1,s_2,s_3)\\
&\quad + O\left(n^{-2}\right)\ . \qquad (2.18)
\end{aligned}$$

Set $s. = \sum_{j=1}^{k} s_j$. Then

$$\begin{aligned}
&B_{1,\cdot} = B_{1,1} - 1\ , \qquad B_{2,2} = 1/s_2\ , \qquad B_{2,2} = 1/s_2\ , \qquad B_{2,2} = s_1\ ,\\
&B_{2,\cdot} = s_1^{-1} + s_2^{-1} = (s_1+s_2)/(s_1 s_2)\ ,\\
&B_{3,1} = (s_2-1)^{-1}\,s_3^{-1}\ , \qquad B_{3,2} = (s_1-1)^{-1}\,s_3^{-1}\ , \qquad B_{3,3} = (s_1-1)^{-1}\,s_2^{-1}\ ,\\
&B_{3,\cdot} = \left\{s_2(s.-2)-s_3\right\}\,(s_1-1)^{-1}\,\langle s_2\rangle_2^{-1} s_3^{-1}\ ,\\
&B_{4,1} = (s_2-2)^{-1}(s_3-1)^{-1}\,s_4^{-1}\ , \qquad B_{4,2} = (s_1-2)^{-1}(s_3-1)^{-1}\,s_4^{-1}\ ,\\
&B_{4,3} = (s_1-2)^{-1}(s_2-1)^{-1}\,s_4^{-1}\ , \qquad B_{4,4} = (s_1-2)^{-1}(s_2-1)^{-1}\,s_3^{-1}\ ,\\
&B_{4,\cdot} = \left\{s.s_3(s_2-2)+s_3(s_2-4s_2+4)-s_2s_4\right\}\left\{(s_1-2)\,\langle s_2-2\rangle_2\,\langle s_3\rangle_2\,s_4\right\}^{-1}\ .
\end{aligned}$$

Also $E_{\mathbf{c}} = c_0^{-2}c_1$, $D_a = s_1^{-1}+s_2^{-1}$, $F_2 = 0$, and

$$\begin{aligned}
&(2.19)\quad \mathbb{E}\left[Y_{n,s}\right] = s^{-1} + n^{-1}E_{\mathbf{c}} + O\left(n^{-2}\right) \qquad \text{for} \quad s>0\ ,\\
&(2.20)\quad \mathbb{E}\left[Y_{n,s_1}Y_{n,s_2}\right] = \left(1-n^{-1}\right)B_{2,\cdot} + n^{-1}E_{\mathbf{c}}D_a + O\left(n^{-2}\right) \qquad \text{if} \quad (2.17) \text{ holds}\ ,\\
&(2.21)\quad \mathrm{Cov}\left(Y_{n,s_1},Y_{n,s_2}\right) = \langle s_1\rangle_2^{-1}s_2^{-1}\left(s_2-n^{-1}s_1\right) + O\left(n^{-2}\right) \qquad \text{if} \quad (2.17) \text{ holds}\ .
\end{aligned}$$

In the case $a\geq 2$, (2.19)–(2.21) hold with $E_{\mathbf{c}}$ replaced by 0. In the case $a\leq 1$, (2.14)–(2.16) with $a_0=a$ give terms $O\left(n^{-2a}\right)$ with the n^{-1} terms disposable if $a\leq 1/2$.

We now investigate what extra terms are needed to make (2.19)–(2.21) depend on c when $a=1$ or 2.

Example 2.2. $\alpha=\beta=1$. Here, we find the coefficients of n^{-2}. By (2.12),

$$\begin{aligned} d_2(\mathbf{s}:\boldsymbol{\psi}) &= \sum_{j=0}^{2} e_{2-j}\left(j-\overline{\psi}_1\right) C_j(\mathbf{s}:\boldsymbol{\psi}) \\ &= e_2\left(-\overline{\psi}_1\right) C_0(\mathbf{s}:\boldsymbol{\psi}) + e_1\left(1-\overline{\psi}_1\right) C_1(\mathbf{s}:\boldsymbol{\psi}) + C_2(\mathbf{s}:\boldsymbol{\psi}) \\ &= C_2(\mathbf{s}:\boldsymbol{\psi}) \qquad \text{if } \ \overline{\psi}_1 = 1 \ \text{ or } \ 2\,. \end{aligned}$$

For $k=1$, $C_2(s:\psi)=C_{2,\psi}(s+1)_{2-\psi}$, where $C_{2,\psi}=\psi c_0^{\psi-4}\left\{c_0c_2+(\psi-3)c_1^2/2\right\}$, so $d_2(s:1)=(s+1)F_{\mathbf{c}}$, where $F_{\mathbf{c}}=c_0^{-3}\left(c_0c_2-c_1^2\right)$, so in (2.19) we may replace $O\left(n^{-2}\right)$ by $n^{-2}(s+1)F_{\mathbf{c}}c_0^{-1}+O\left(n^{-3}\right)$. For $k=2$,

$$\begin{aligned} C_2(\mathbf{s}:\mathbf{1}) &= \sum\Big\{C_{i,1}C_{j,1}B(\mathbf{s}:0,j-1):\ i+j=2\Big\} \\ &= C_{0,1}C_{2,1}\big\{B(\mathbf{s}:0,1)+B(\mathbf{s}:0,-1)\big\}+C_{1,1}^2B(\mathbf{s}:0,0)\ , \end{aligned}$$

where $B(\mathbf{s}:0,\lambda)=b(s_1-s_2,s_2+1:\lambda)=\pi_{\mathbf{s}}(\lambda)$, so $d_2(\mathbf{s}:\mathbf{1})=C_2(\mathbf{s}:\mathbf{1})-D_{2,\mathbf{s}}H_{\mathbf{c}}+c_0^{-2}c_1^2$, where $D_{2,\mathbf{s}}=(s_2+1)(s_1+1)^{-1}+s_1s_2^{-1}$, $H_{\mathbf{c}}=c_0^{-2}\left(c_0c_2-c_1^2\right)$ and in (2.20) we may replace $O\left(n^{-2}\right)$ by $n^{-2}d_2(\mathbf{s}:\mathbf{1})\,c_0^{-2}+O\left(n^{-3}\right)$. Upon simplifying this gives

$$\operatorname{Cov}\left(Y_{n,s_1},Y_{n,s_2}\right) = \langle s_1\rangle_2^{-1}s_2^{-1}\left(1-n^{-1}s_1\right)-c_0^{-2}H_{\mathbf{c}}F_{3,\mathbf{s}}\,n^{-2}+O\left(n^{-2}\right)\ ,$$

where $F_{3,\mathbf{s}}=(s_2+1)/\langle s_1\rangle_2+s_2^{-1}$.

Example 2.3. $\alpha=1$, $\beta=2$. So, $a=2$, $\lambda=1$, $\boldsymbol{\psi}=\boldsymbol{\theta}$. By (2.12),

$$\begin{aligned} d_2(\mathbf{s}:\boldsymbol{\psi}) &= \sum_{j=0}^{1} e_{2-2j}\left(2j-\overline{\psi}_1\right) C_j(\mathbf{s}:\boldsymbol{\psi}) \\ &= e_2\left(-\overline{\psi}_1\right) C_0(\mathbf{s}:\boldsymbol{\psi}) + C_1(\mathbf{s}:\boldsymbol{\psi}) \\ &= C_1(\mathbf{s}:\boldsymbol{\psi}) \qquad \text{if } \ \overline{\psi}_1 = 0,\ 1 \ \text{ or } \ 2\,. \end{aligned}$$

For $k=1$,

$$C_1(s:\psi) = \psi\, c_0^{\psi-3}c_1\langle s\rangle_{\psi-2}^{-1} = \begin{cases} c_0^{-2}c_1(s+1), & \text{if } \psi=1\,, \\ 2\,c_0^{-1}c_1\,, & \text{if } \psi=2\,, \end{cases}$$

so $\mathbb{E}\left[Y_{n,s}\right]=s^{-1}+c_0^{-3}c_1(s+1)n^{-2}+O\left(n^{-3}\right)$ for $s>0$. For $k=2$, $C_1(\mathbf{s}:\mathbf{1})=c_0^{-1}c_1D_{2,\mathbf{s}}$ for $D_{2,\mathbf{s}}$ above, so

$$\mathbb{E}\left[Y_{n,s_1}Y_{n,s_2}\right] = \left(1-n^{-1}\right)(s_1-1)^{-1}s_2^{-1}+n^{-2}c_0^{-3}c_1D_{2,\mathbf{s}}+O\left(n^{-3}\right)$$

and

$$\operatorname{Cov}\left(Y_{n,s_1},Y_{n,s_2}\right) = \langle s_1\rangle_2^{-1}s_2^{-1}\left(1-n^{-1}s_1\right)-n^{-2}c_0^{-3}c_1F_{3,\mathbf{s}}+O\left(n^{-3}\right)\ .$$

3. EXAMPLES

Example 3.1. For Student's t distribution, $X = t_N$ has density function

$$\left(1 + x^2/N\right)^{-\gamma} g_N \;=\; \sum_{i=0}^{\infty} d_i\, x^{-2\gamma-2i}\;,$$

where $\gamma = (N+1)/2$, $g_N = \Gamma(\gamma)/\left\{\sqrt{N\pi}\;\Gamma(N/2)\right\}$ and $d_i = \binom{-\gamma}{i} N^{\gamma+i} g_N$. So, (1.6) holds with $\alpha = N$, $\beta = 2$ and $c_i = d_i/(N+2i)$. In particular,

$$\begin{aligned}
c_0 &= N^{\gamma-1} g_N\;,\\
c_1 &= -\gamma N^{\gamma+1}(N+2)^{-1} g_N \;=\; -N^{\gamma+1}(N+1)\,(N+2)^{-1}\,g_N/2\;,\\
c_2 &= (\gamma)_2\, N^{\gamma+2}(N+4)^{-1} g_N/2\;,\\
c_3 &= -(\gamma)_3\, N^{\gamma+3} g_N (N+6)^{-1}/6\;,
\end{aligned}$$

and so on. So, $a = 2/N$ and (2.12) gives an expression in powers of $n^{-a/2}$ if N is odd or n^{-a} if N is even. The first term in (2.12) to involve c_1, not just c_0, is the coefficient of n^{-a}.

Putting $N = 1$ we obtain

Example 3.2. For the Cauchy distribution, (1.6) holds with $\alpha = 1$, $\beta = 2$ and $c_i = (-1)^i(2i+1)^{-1}\pi^{-1}$. So, $a = 2$, $\psi = \theta$, $C_{0,\psi} = \pi^{-\psi}$, $C_{1,\psi} = -\psi\,\pi^{2-\psi}/3$, $C_{2,\psi} = \psi\,\pi^{4-\psi}\left\{1/5 + (\psi-5)/a\right\}$ and $C_{3,\psi} = -\psi\,\pi^{6-\psi}\left\{1/105 - 2\psi/15 + (\psi+1)_2/162\right\}$. By Example 2.3, $Y_{n,s} = (\pi/n)\,X_{n,n-s}$ satisfies

$$\mathbb{E}\left[Y_{n,s}\right] \;=\; s^{-1} - n^{-2}\pi^2(s+1) + O\left(n^{-3}\right) \tag{3.1}$$

for $s > 0$ and when (2.17) holds

$$\mathbb{E}\left[Y_{n,s_1}Y_{n,s_2}\right] \;=\; \left(1-n^{-1}\right)(s_1-1)^{-1}\,s_2^{-1} - n^{-2}\pi^2\,D_{2,\mathbf{s}}/3 + O\left(n^{-3}\right) \tag{3.2}$$

for $D_{2,\mathbf{s}} = (s_2+1)\,/\,(s_1+1) + s_1/s_2$ and

$$\mathrm{Cov}\left(Y_{n,s_1}, Y_{n,s_2}\right) \;=\; \langle s_1\rangle_2^{-1} s_2^{-1}\left(1 - n^{-1}s_1\right) + n^{-2}\pi^2\,F_{3,\mathbf{s}}/3 + O\left(n^{-3}\right)$$

for $F_{3,\mathbf{s}} = (s_2+1)\,/\langle s_1\rangle_2 + s_2^{-1}$. Page 274 of Hall [7] gave the first term in (3.1) and (3.2) when $s_1 = s_2$ but his version of (3.2) for $s_1 > s_2$ replaces $(s_1-1)^{-1}\,s_2^{-1}$ and $D_{2,\mathbf{s}}$ by complicated expressions each with $s_1 - s_2$ terms. The joint order of order three for $\{Y_{n,s_i}, 1 \le i \le 3\}$ is given by (2.18). Hall points out that $F^{-1}(u) = \cot(\pi - \pi u)$, so $F^{-1}(u) = \sum_{i=0}^{\infty}(1-u)^{2i-1}C_{i,1}$, where $C_{i,1} = \left(-4\pi^2\right)^i\pi^{-1}B_{2,i}/(2i)!$.

Example 3.3. Consider the F distribution. For $N, M \geq 1$, set $\nu = M/N$, $\gamma = (M+N)/2$ and $g_{M,N} = \nu^{M/2}/B\,(M/2, N/2)$. Then $X = F_{M,N}$ has density function

$$x^{M/2}\,(1+\nu x)^{-\gamma}\, g_{M,N} \;=\; \nu^{-\gamma}\, x^{-N/2}\left(1+\nu^{-1}x^{-1}\right)^{-\gamma} g_{M,N} \;=\; \sum_{i=0}^{\infty} d_i\, x^{-N/2-i}\ ,$$

where $d_i = h_{M,N}\binom{-\gamma}{i}\nu^i$ and $h_{M,N} = g_{M,N}\,\nu^{-\gamma} = \nu^{-N/2}/B(M/2, N/2)$. So, for $N > 2$, (2.1) holds with $\alpha = N/2 - 1$, $\beta = 1$ and $c_i = d_i/(N/2 + i - 1)$. If $N = 4$ then $\alpha = 1$ and Examples 2.1–2.2 apply. Otherwise (2.13)–(2.15) give $\mathbb{E}\,[Y_{n,\mathrm{s}}]$, $\mathbb{E}\,[Y_{n,s_1}Y_{n,s_2}]$ and $\mathrm{Cov}\,(Y_{n,s_1}, Y_{n,s_2})$ to $O\left(n^{-2a_0}\right)$, where $Y_{n,s} = X_{n,n-s}/\left(nc_0\right)\lambda$, $\lambda = 1/\alpha$, $a = 2/(N-2)$, $a_0 = \min(a, 1) = a$ if $N \geq 4$ and $a_0 = \min(a, 1) = 1$ if $N < 4$.

Example 3.4. Consider the stable laws. Page 549 of Feller [5] proves that the general stable law of index $\alpha \in (0, 1)$ has density function

$$\sum_{k=1}^{\infty} |x|^{-1-\alpha k} a_k(\alpha, \gamma)\ ,$$

where $a_k(\alpha, \gamma) = (1/\pi)\,\Gamma(k\alpha + 1)\left\{(-1)^k/k!\right\}\,\sin\{k\pi(\gamma - \alpha)/2\}$ and $|\gamma| \leq \alpha$. So, for $x > 0$ its distribution function F satisfies (2.1) with $\beta = \alpha$ and $c_i = a_{i+1}(\alpha, \gamma)\,\gamma^{-1}(i+1)^{-1}$. Since $a = 1$ the first two moments of $Y_{n,s} = X_{n,n-s}/\left(nc_0\right)^{\lambda}$, where $\lambda = 1/\alpha$ are $O\left(n^{-2}\right)$ by (2.13)–(2.15).

Example 3.5. Finally, consider the second extreme value distribution. Suppose $F(x) = \exp\left(-x^{-\alpha}\right)$ for $x > 0$, where $\alpha > 0$. Then (1.6) holds with $\beta = \alpha$ and $c_i = (-1)^i/(i+1)!$. Since $a = 1$ the first two moments of $Y_{n,s} = X_{n,n-s}/n^{1/\alpha}$ are given to $O\left(n^{-2}\right)$ by (2.13)–(2.15).

APPENDIX: AN INVERSION THEOREM

Given $x_j = y_j/j!$ for $j \geq 1$ set

$$S = \widehat{S}(t, \mathbf{x}) = \sum_{j=1}^{\infty} x_j\, t^j = S(t, \mathbf{y}) = \sum_{j=1}^{\infty} y_j\, t^j/j!\,. \tag{A.1}$$

The partial ordinary and exponential Bell polynomials $\widehat{B}_{r,i}(\mathbf{x})$ and $B_{r,i}(\mathbf{y})$ are defined for $r = 0, 1, ...$ by

$$S^i = \sum_{r=i}^{\infty} t^r\, \widehat{B}_{r,i}(\mathbf{x}) = i! \sum_{r=i}^{\infty} t^r\, B_{r,i}(\mathbf{y})/r!\,.$$

So, $\widehat{B}_{r,0}(\mathbf{x}) = B_{r,0}(\mathbf{y}) = I(r = 0)$, $\widehat{B}_{r,i}(\lambda\mathbf{x}) = \lambda^i \widehat{B}_{r,i}(\mathbf{x})$ and $B_{r,i}(\lambda\mathbf{y}) = \lambda^i B_{r,i}(\mathbf{y})$. They are tabled on pages 307–309 of Comtet [2] for $r \leq 10$ and 12. Note that

$$(1 + \lambda S)^{\alpha} = \sum_{r=0}^{\infty} t^r\, \widehat{C}_r = \sum_{r=0}^{\infty} t^r\, C_r/r!\,, \tag{A.2}$$

where

$$\widehat{C}_r = \widehat{C}_r\,(\alpha, \lambda, \mathbf{x}) = \sum_{i=0}^{r} \widehat{B}_{r,i}(\mathbf{x}) \binom{\alpha}{i} \lambda^i \tag{A.3}$$

and

$$C_r = C_r\,(\alpha, \lambda, \mathbf{y}) = \sum_{i=0}^{r} B_{r,i}(\mathbf{y})\, \langle\alpha\rangle_i\, \lambda^i\,.$$

So, $\widehat{C}_0 = 1$, $\widehat{C}_1 = \alpha\lambda x_1$, $\widehat{C}_2 = \alpha\lambda x_2 + \langle\alpha\rangle_2 \lambda^2 x_1^2/2$, $\widehat{C}_3 = \alpha\lambda x_3 + \langle\alpha\rangle_2 \lambda^2 x_1 x_2 + \langle\alpha\rangle_3 \lambda^3 x_1^3/6$ and $C_0 = 1$, $C_1 = \alpha\lambda y_1$, $C_2 = \alpha\lambda y_2 + \langle\alpha\rangle_2 \lambda^2 y_1^2$. Similarly,

$$\log(1 + \lambda S) = \sum_{r=1}^{\infty} t^r\, \widehat{D}_r = \sum_{r=1}^{\infty} t^r\, D_r/r!$$

and

$$\exp(\lambda S) = 1 + \sum_{r=1}^{\infty} t^r\, \widehat{B}_r = 1 + \sum_{r=1}^{\infty} t^r\, B_r/r!\,,$$

where

$$\widehat{D}_r = \widehat{D}_r(\lambda, \mathbf{x}) = -\sum_{i=1}^{r} \widehat{B}_{r,i}(\mathbf{x})\, (-\lambda)^i/i!\,,$$

$$D_r = D_r(\lambda, \mathbf{y}) = -\sum_{i=1}^{r} B_{r,i}(\mathbf{y})\, (-\lambda)^i/(i-1)!\,,$$

$$\widehat{B}_r = \widehat{B}_r(\lambda, \mathbf{x}) = \sum_{i=1}^{r} \widehat{B}_{r,i}(\mathbf{x})\, \lambda^i / i!$$

and

$$B_r = B_r(\lambda, \mathbf{y}) = \sum_{i=1}^{r} B_{r,i}(\mathbf{y})\, \lambda^i \,.$$

Here, $\widehat{B}_r(1, \mathbf{x})$ and $B_r(1, \mathbf{y})$ are known as the *complete* ordinary and exponential Bell polynomials. If $x_j = y_j = 0$ for j even, then $S = t^{-1} \sum_{j=1}^{\infty} X_j t^{2j}$, where $X_j = x_{2j-1}$, so

$$S^i = t^{-i} \sum_{r=i}^{\infty} t^{2r} \widehat{B}_{r,i}(\mathbf{X}) \qquad \text{and} \qquad \exp(\lambda S) = 1 + \sum_{k=1}^{\infty} t^k \widehat{B}_k \,,$$

where

$$\widehat{B}_k = \sum \left\{ \widehat{B}_{r,i}(\mathbf{X}) \lambda^i / i! \colon\ i = 2r - k,\ k/2 < r \le k \right\} .$$

The following derives from Lagrange's inversion formula.

Theorem A.1. *Let k be a positive integer and a any real number. Suppose*

$$v/u = \sum_{i=0}^{\infty} x_i\, u^{ia} = \sum_{i=0}^{\infty} y_i\, v^{ia} / i!$$

with $x_0 \neq 0$. Then

$$(u/v)^k = \sum_{i=0}^{\infty} x_i^*\, v^{ia} = \sum_{i=0}^{\infty} y_i^* v^{ia} / (ia)! \,,$$

where $x_i^* = x_i^*\,(a, k, \mathbf{x})$ and $y_i^* = y_i^*\,(a, k, \mathbf{y})$ are given by

$$\text{(A.4)} \qquad x_i^* = k\, n^{-1} \widehat{C}_i\,(-n, 1/x_0, \mathbf{x}) = k\, x_0^{-n} \sum_{j=0}^{i} (n+1)_{j-1} \widehat{B}_{i,j}\,(\mathbf{x})\,(-x_0)^{-j} / j!$$

and

$$\text{(A.5)} \qquad y_i^* = k\, n^{-1} C_i\,(-n, 1/y_0, \mathbf{y}) = k\, y_0^{-n} \sum_{j=0}^{i} (n+1)_{j-1} B_{i,j}(\mathbf{y})\,(-y_0)^{-j} \,,$$

respectively, where $n = k + a\,i$.

Proof: u/v has a power series in v^a so that $(u/v)^k$ does also. A little work shows that (A.4)–(A.5) are correct for $i = 0, 1, 2, 3$ and so by induction that $x_i^* x_0^{ia}$ and $y_i^* y_0^{ia}$ are polynomials in a of degree $i - 1$. Hence, (A.4)–(A.5) will hold true for all a if they hold true for all positive integers a. Suppose then a is a positive integer. Since $v/u = x_0\left(1 + x_0^{-1}S\right)$ for $S = \widehat{S}\,(u^a, \mathbf{x}) = S\,(u^a, \mathbf{y})$, the coefficient of u^{ai} in $(v/u)^{-n}$ is $x_0^{-n}\widehat{C}_i\,(-n, 1/x_0, \mathbf{x}) = y_0^{-n}C_i\,(-n, 1/y_0, \mathbf{y})\,/(n-k)!$. Now set $n = k + a\,i$ and apply Theorem A in page 148 of Comtet [2] to $v = f(u) = \sum_{i=0}^{\infty} x_i u^{1+ai}$. $\square$

Theorem F in page 15 of Comtet [2] proves (A.4) for the case $k = 1$ and a a positive integer.

ACKNOWLEDGMENTS

The authors would like to thank the Editor and the referee for careful reading and comments which greatly improved the paper.

REFERENCES

[1] Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions*, National Bureau of Standards, Washington, D.C.

[2] Comtet, L. (1974). *Advanced Combinatorics*, Reidel, Dordrecht.

[3] David, F. N. and Johnson, N. L. (1954). Statistical treatment of censored data. Part I: Fundamental formulae, *Biometrika*, **41**, 225–231.

[4] Downey, P. J. (1990). Distribution-free bounds on the expectation of the maximum with scheduling applications, *Operations Research Letters*, **9**, 189–201.

[5] Feller, W. (1966). *An Introduction to Probability Theory and Its Applications*, volume 2, John Wiley and Sons, New York.

[6] Gomes, M. I. and Guillou, A. (2014). Extreme value theory and statistics of univariate extremes: A review, *International Statistical Review*, doi: 10.1111/insr.12058.

[7] Hall, P. (1978). Some asymptotic expansions of moments of order statistics, *Stochastic Processes and Their Applications*, **7**, 265–275.

[8] Hill, B. (1975). A simple general approach to inference about the tail of a distribution, *Annals of Statistics*, **3**, 1163–1173.

[9] Hill, T. P. and Spruill, M. C. (1994). On the relationship between convergence in distribution and convergence of expected extremes, *Proceedings of the American Mathematical Society*, **121**, 1235–1243.

[10] Hüsler, J.; Piterbarg, V. and Seleznjev, O. (2003). On convergence of the uniform norms for Gaussian processes and linear approximation problems, *Annals of Applied Probability*, **13**, 1615–1653.

[11] James, G. S. and Mayne, A. J. (1962). Cumulants of functions of random variables, *Sankhyā*, A, **24**, 47–54.

[12] McCord, J. R. (1964). On asymptotic moments of extreme statistics, *Annals of Mathematical Statistics*, **64**, 1738–1745.

[13] Nair, K. A. (1981). Asymptotic distribution and moments of normal extremes, *Annals of Probability*, **9**, 150–153.

[14] Novak, S. Y. and Utev, S. A. (1990). Asymptotics of the distribution of the ratio of sums of random variables, *Siberian Mathematical Journal*, **31**, 781–788.

[15] Pickands, J. (1968). Moment convergence of sample extremes, *Annals of Mathematical Statistics*, **39**, 881–889.

[16] Ramachandran, G. (1984). Approximate values for the moments of extreme order statistics in large samples. In: *Statistical Extremes and Applications* (Vimeiro, 1983), pp. 563–578, NATO Advanced Science Institutes Series C: Mathematical and Physical Sciences, volume 131, Reidel, Dordrecht.

[17] Stuart, A. and Ord, J. K. (1987). *Kendall's Advanced Theory of Statistics*, fifth edition, volume 1, Griffin, London.

[18] Withers, C. S. and Nadarajah, S. (2014). Asymptotic multivariate distributions and moments of extremes, *Technical Report*, Applied Mathematics Group, Industrial Research Ltd., Lower Hutt, New Zealand.

REVSTAT – Statistical Journal
Volume 15, Number 1, January 2017, 45–64

COMPARISON OF THE PREDICTIVE VALUES OF MULTIPLE BINARY DIAGNOSTIC TESTS IN THE PRESENCE OF IGNORABLE MISSING DATA

Authors: Ana Eugenia Marín-Jiménez
– Statistics (Biostatistics), Faculty of Education, Economy and Technology, University of Granada, Ceuta, Spain
anamarin@ugr.es

José Antonio Roldán-Nofuentes
– Statistics (Biostatistics), School of Medicine, University of Granada, Granada, Spain
jaroldan@ugr.es

Received: August 2013 Revised: April 2015 Accepted: May 2015

Abstract:

- The comparison of the predictive values of binary diagnostic tests is an important topic in the study of statistical methods applied to medical diagnosis. In this article, we study a global hypothesis test to simultaneously compare the predictive values of multiple binary diagnostic tests in the presence of ignorable missing data. The global hypothesis test deduced is based on the chi-squared distribution. Simulation experiments were carried out to study the type I error probability and the power of global hypothesis test and of other alternative methods when comparing the predictive values of two and three binary diagnostic tests respectively.

Key-Words:

- *global hypothesis test; predictive values; multiple comparisons; chi-squared distribution; ignorable missing data.*

AMS Subject Classification:

- 62P10, 62H15.

1. INTRODUCTION

A diagnostic test is a medical test that is applied to an individual in order to determine the presence or absence of a disease. When the diagnostic test can only give two possible results (positive, indicating the provisional presence of the disease, or negative, indicating the provisional absence of the disease) the diagnostic test is called a binary diagnostic test (BDT) and it is used very frequently in clinical practice. A stress test for the diagnosis of coronary disease or a mammogram to diagnose breast cancer are two examples of *BDTs*. The most common parameters to assess the accuracy or performance of a BDT are sensitivity (Se) and specificity (Sp). Other commonly used parameters to assess the performance of a BDT are predictive values (PVs). The positive predictive value (PPV) of a BDT is the probability of an individual having the disease given that the result of the BDT is positive and the negative predictive value (NPV) is the probability of an individual not having the disease given that the result of the BDT is negative. The predictive values (PVs) are a measure of clinical accuracy of the *BDT*, and they depend on the sensitivity and the specificity of the *BDT* and on the prevalence of the disease (p). Applying Bayes Theorem, the PVs are calculated as

$$PPV = \frac{p \times Se}{p \times Se + (1-p) \times (1-Sp)}$$

(1.1) and

$$NPV = \frac{(1-p) \times Sp}{p \times (1-Se) + (1-p) \times Sp} .$$

In the study of statistical methods for the diagnosis of diseases, comparison of the accuracy or the performance of two diagnostic tests is a topic of particular importance. In paired designs (*i.e.* when the two *BDTs* and the gold standard are applied to all of the individuals in a random sample), comparison of the PVs of two *BDTs* in relation to the same gold standard has been the subject of several studies in the statistical literature [1, 2, 3, 4]. In all of them, the comparison of the two $PPVs$ and the comparison of the two $NPVs$ is carried out independently. Roldán-Nofuentes *et al.* [5] showed that the PVs of two (or more) *BDTs* are correlated and they studied a global hypothesis test based on the chi-squared distribution to simultaneously compare the PVs of two or more *BDTs* in relation to the same gold standard. In all of these studies, the disease status of all of the patients is known, as well as the results of the two diagnostic tests. This situation is also known as 'complete verification' (because the gold standard is applied to all of the individuals in the sample). Poleto *et al.* [6] studied the comparison of the predictive values of two *BDTs* when for some individuals we do not know the results of one of the two *BDTs*. Furthermore, in clinical practice it is common for the gold standard not to be applied to all of the individuals in the sample, thus leading to the problem known as partial disease verification [7, 8, 9].

Therefore, the disease status (if the disease is present or absent) is unknown for a subset of individuals in the sample. In this situation, Roldán-Nofuentes *et al.* [10, 11] studied the comparison of the *PPVs* and of the *NPVs* of two *BDTs*. Nevertheless, in these two studies they did not consider the dependence that exists between the *PVs* of the diagnostic tests. This is the essence of our article, to study a global hypothesis test that allows us to jointly compare the *PVs* of two (or more) *BDTs* in the presence of ignorable missing data. In this article, we study a global hypothesis test to simultaneously compare the predictive values of two or more *BDTs* when, in the presence of partial disease verification, the missing data mechanism is ignorable. In Section 2, we propose a global hypothesis test, and other alternative methods, to simultaneously compare the *PVs* of multiple *BDTs*. In Section 3, Monte Carlo simulation experiments are carried out in order to study the type I error probability and the power of the global hypothesis test (and of the alternative methods) when comparing the *PVs* of two and of three *BDTs* respectively. In Section 4, the method proposed is applied to two examples, and in Section 5 the results obtained are discussed.

2. THE MODEL

Let us consider J *BDTs* ($J \geqslant 2$) that are applied independently to the same random sample of size n extracted from a population that has a determined prevalence of the disease (p). Moreover, let us consider that the gold standard has not been applied to all of the individuals in the random sample. In this situation, the J diagnostic tests are applied to all of the individuals in the sample whilst the gold standard is only applied to a subset of them. Therefore, the results of the J diagnostic tests are known by all of the individuals in the sample, whereas the result of the gold standard (*i.e.* the disease status) is only unknown to a subset of them. Let T_j, V and D be the random binary variables defined as: T_j which models the result of the j-th BDT ($j = 1, ..., J$), so that $T_j = 1$ when the test result is positive and $T_j = 0$ when the result is negative; V models the verification process, $V = 1$ when the individual is verified with the gold standard and $V = 0$ when the individual is not verified; and D models the result of the gold standard, $D = 1$ when the individual has the disease and $D = 0$ when the individual does not. Let $Se_j = P\,(T_j = 1 \,|\, D = 1)$, $Sp_j = P\,(T_j = 0 \,|\, D = 0)$, $PPV_j = P\,(D = 1 \,|\, T_j = 1)$ and $NPV_j = P\,(D = 0 \,|\, T_j = 0)$ be the sensitivity, the specificity, the positive predictive value and the negative predictive value of the j-th BDT respectively. Let the observed frequencies be: $s_{i_1,...,i_J}$ is the number of patients verified in which $T_1 = i_1$, $T_2 = i_2$, ..., $T_J = i_J$ and $D = 1$; $r_{i_1,...,i_J}$ is the number of patients verified in which $T_1 = i_1$, $T_2 = i_2$, ..., $T_J = i_J$ and $D = 0$; and $u_{i_1,...,i_J}$ is the number of patients not verified in which $T_1 = i_1$, $T_2 = i_2$, ..., $T_J = i_J$ with $i_j = 0, 1$ and $j = 1, ..., J$. Let $n_{i_1,...,i_J} = s_{i_1,...,i_J} + r_{i_1,...,i_J} + u_{i_1,...,i_J}$ and

$n = \sum_{i_1,\ldots,i_J=0}^{1} n_{i_1,\ldots,i_J}$. As only a subset of individuals in the sample have their disease status verified with the gold standard, the verification probabilities $(\lambda_{k,i_1,\ldots,i_J})$ are defined as the probability of selecting an individual for whom $D = k$, $T_1 = i_1$, $T_2 = i_2$, ..., $T_J = i_J$ with $k, i_j = 0, 1$, $j = 1, \ldots, J$, to verify his or her disease status *i.e.*

$$\lambda_{k,i_1,\ldots,i_J} = P\Big(V=1 \,\Big|\, D=k,\, T_1=i_1,\, T_2=i_2,\, \ldots,\, T_J=i_J\Big)\,.$$

Assuming that the verification process with the gold standard only depends on the results of the J *BDTs* and does not depend on the disease status, then the missing data mechanism is missing at random (MAR) [12]. Assuming also that the parameters of the data model and the parameters of the missingness mechanism are distinct, then the missing data mechanism is ignorable [13]. Under this model, the verification probabilities are

$$\lambda_{k,i_1,\ldots,i_J} = \lambda_{i_1,\ldots,i_J} = P\Big(V=1 \,\Big|\, T_1=i_1,\, T_2=i_2,\, \ldots,\, T_J=i_J\Big)\,,$$

and all of the parameters can be estimated applying the maximum likelihood method.

2.1. Maximum likelihood estimators of the *PVs*

As the J *BDTs* are applied to all of the n individuals in the random sample and the gold standard is only applied to a subset of them, the frequencies observed $r_{i_1,\ldots,i_J}$, $s_{i_1,\ldots,i_J}$ and $u_{i_1,\ldots,i_J}$ with $i_j = 0, 1$ and $j = 1, \ldots, J$, which can be written in the form of a 3×2^J table in which the sample of size n has been set, are the realization of a multinomial distribution whose probabilities are

$$\phi_{i_1,\ldots,i_J} = P\Big(V=1,\, D=1,\, T_1=i_1,\, T_2=i_2,\, \ldots,\, T_J=i_J\Big)\,,$$

$$\varphi_{i_1,\ldots,i_J} = P\Big(V=1,\, D=0,\, T_1=i_1,\, T_2=i_2,\, \ldots,\, T_J=i_J\Big)$$

and

$$\gamma_{i_1,\ldots,i_J} = P\Big(V=0,\, T_1=i_1,\, T_2=i_2,\, \ldots,\, T_J=i_J\Big)\,.$$

Let $\boldsymbol{\omega} = (\phi_{1,\ldots,1}, \ldots, \phi_{0,\ldots,0}, \varphi_{1,\ldots,1}, \ldots, \varphi_{0,\ldots,0}, \gamma_{1,\ldots,1}, \ldots, \gamma_{0,\ldots,0})^{\mathrm{T}}$ be a vector sized $\left(3 \cdot 2^J\right)$ whose components are the probabilities of multinomial distribution and $\eta_{i_1,\ldots,i_J} = \phi_{i_1,\ldots,i_J} + \varphi_{i_1,\ldots,i_J} + \gamma_{i_1,\ldots,i_J}$. Assuming that the missing data mechanism is ignorable, the *PVs* of the j-th *BDT* are written in terms of the parameters of the vector $\boldsymbol{\omega}$ and of the verification probabilities as

$$PPV_j = \frac{\sum\limits_{i_1,\ldots,i_J=0;\, i_j=1}^{1} \phi_{i_1,\ldots,i_J}\, \lambda^{-1}_{i_1,\ldots,i_J}}{\sum\limits_{i_1,\ldots,i_J=0;\, i_j=1}^{1} \eta_{i_1,\ldots,i_J}}$$

(2.1) and

$$NPV_j = \frac{\sum\limits_{i_1,\ldots,i_J=0;\, i_j=0}^{1} \varphi_{i_1,\ldots,i_J}\, \lambda^{-1}_{i_1,\ldots,i_J}}{\sum\limits_{i_1,\ldots,i_J=0;\, i_j=0}^{1} \eta_{i_1,\ldots,i_J}} ,$$

where $\lambda_{i_1,\ldots,i_J} = (\phi_{i_1,\ldots,i_J} + \varphi_{i_1,\ldots,i_J})/\eta_{i_1,\ldots,i_J}$ are the verification probabilities. Therefore, in equations (2.1) we can observe the dependence of the *PVs* of the verification process subject to the MAR assumption. In this situation the logarithm of the likelihood function is

$$\begin{aligned} l \;=\; & \sum_{i_1,\ldots,i_J=0}^{1} s_{i_1,\ldots,i_J} \log\left(\phi_{i_1,\ldots,i_J}\right) + \sum_{i_1,\ldots,i_J=0}^{1} r_{i_1,\ldots,i_J} \log\left(\varphi_{i_1,\ldots,i_J}\right) \\ & + \sum_{i_1,\ldots,i_J=0}^{1} u_{i_1,\ldots,i_J} \log\left(\gamma_{i_1,\ldots,i_J}\right) , \end{aligned}$$

so that maximizing this function, the maximum likelihood estimators (MLEs) of $\phi_{i_1,\ldots,i_J}$, $\varphi_{i_1,\ldots,i_J}$ and $\gamma_{i_1,\ldots,i_J}$ are the estimators of multinomial proportions [14], *i.e.*

$$\text{(2.2)} \qquad \hat{\phi}_{i_1,\ldots,i_J} = \frac{s_{i_1,\ldots,i_J}}{n} , \qquad \hat{\varphi}_{i_1,\ldots,i_J} = \frac{r_{i_1,\ldots,i_J}}{n} \quad \text{and} \quad \hat{\gamma}_{i_1,\ldots,i_J} = \frac{u_{i_1,\ldots,i_J}}{n} ,$$

and the *MLE* of $\eta_{i_1,\ldots,i_J}$ is $\hat{\eta}_{i_1,\ldots,i_J} = n_{i_1,\ldots,i_J}/n$. Substituting in equations (2.1) the parameters with their respective *MLEs* given in equations (2.2), the *MLEs* of the *PVs* of the j-th *BDT* are

$$\widehat{PPV_j} = \frac{\sum\limits_{i_1,\ldots i_J=0;\, i_j=1}^{1} \frac{s_{i_1,\ldots i_J}\, n_{i_1,\ldots i_J}}{s_{i_1,\ldots i_J} + r_{i_1,\ldots i_J}}}{\sum\limits_{i_1,\ldots i_J=0;\, i_j=1}^{1} n_{i_1,\ldots i_J}}$$

and

$$\widehat{NPV_j} = \frac{\sum\limits_{i_1,\ldots i_J=0;\, i_j=0}^{1} \frac{r_{i_1,\ldots i_J}\, n_{i_1,\ldots i_J}}{s_{i_1,\ldots i_J} + r_{i_1,\ldots i_J}}}{\sum\limits_{i_1,\ldots i_J=0;\, i_j=0}^{1} n_{i_1,\ldots i_J}} .$$

Once we have obtained the *MLEs* of the *PVs* of the *J BDTs*, we then estimate their variances-covariances.

2.2. Estimation of the variances-covariances of the *PVs*

As the vector $\boldsymbol{\omega}$ is the vector of probabilities of a multinomial distribution, the variance-covariance matrix of $\hat{\boldsymbol{\omega}}$ is $\sum_{\hat{\boldsymbol{\omega}}} = \left\{\text{diag}\left(\boldsymbol{\omega}\right) - \boldsymbol{\omega}^{\text{T}}\boldsymbol{\omega}\right\}/n$. Let $\boldsymbol{\tau} = \left(PPV_1, ..., PPV_J, NPV_1, ..., NPV_J\right)^{\text{T}}$ be a vector sized $2J$ whose components are the *PVs* of the *J BDTs*, and let $\hat{\boldsymbol{\tau}}$ be the *MLE* of $\boldsymbol{\tau}$. As $\boldsymbol{\tau}$ is a function of the components of the vector $\boldsymbol{\omega}$, applying the delta method [15] the asymptotic variance-covariance matrix of $\hat{\boldsymbol{\tau}}$ is

$$\sum\nolimits_{\hat{\boldsymbol{\tau}}} = \left(\frac{\partial\boldsymbol{\tau}}{\partial\boldsymbol{\omega}}\right)\sum\nolimits_{\hat{\boldsymbol{\omega}}}\left(\frac{\partial\boldsymbol{\tau}}{\partial\boldsymbol{\omega}}\right)^{\text{T}}.$$

Substituting in the previous expression each parameter with its corresponding *MLE*, we obtain the estimated asymptotic variances-covariances of the estimators of the *PVs* of the *J BDTs*.

Moreover, the asymptotic variances-covariances of $\hat{\boldsymbol{\tau}}$ can also be estimated through bootstrap [16], generating, from the random sample of size n, B samples with replacement and from these B samples asymptotic variance-covariance matrix of $\hat{\boldsymbol{\tau}}$ is estimated.

Once we have obtained the *MLEs* of the *PVs* and their estimated asymptotic variances-covariances, it is possible to solve the global hypothesis test to simultaneously compare the *PVs* of the *J BDTs*.

2.3. Global hypothesis test

The global hypothesis test to simultaneously compare the *PVs* of *J BDTs* is

$$\begin{aligned} &H_0\colon\quad PPV_1 = PPV_2 = \cdots = PPV_J \quad \text{and} \quad NPV_1 = NPV_2 = \cdots = NPV_J\ , \\ &H_1\colon\quad \text{at least one equality is not true}\ , \end{aligned}$$

which is equivalent to the hypothesis test

$$H_0\colon\ \mathbf{A}\boldsymbol{\tau} = 0 \qquad \text{vs} \qquad H_1\colon\ \mathbf{A}\boldsymbol{\tau} \neq 0\ , \tag{2.3}$$

where $\mathbf{A}$ is a full rank matrix sized $2(J-1) \times 2J$ whose elements are known constants. For two *BDTs* ($J = 2$) the matrix $\mathbf{A}$ is $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & -1 \end{pmatrix}$, and for $J = 3$ this matrix is $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}$, where $\otimes$ is the Kronecker product. As the vector $\hat{\boldsymbol{\tau}}$ is asymptotically distributed according to a normal multivariate distribution, *i.e.* $\hat{\boldsymbol{\tau}} \xrightarrow[n\to\infty]{} N\left(\boldsymbol{\tau}, \boldsymbol{\Sigma}_{\boldsymbol{\tau}}\right)$, the Wald statistic for the global hypothesis test (2.3) is

$$Q^2 = \hat{\boldsymbol{\tau}}^{\mathrm{T}} \mathbf{A}^{\mathrm{T}} \left(\mathbf{A} \hat{\sum}_{\hat{\boldsymbol{\tau}}} \mathbf{A}^{\mathrm{T}} \right)^{-1} \mathbf{A} \hat{\boldsymbol{\tau}} \,, \tag{2.4}$$

which is asymptotically distributed according to a chi-squared distribution with $2\,(J-1)$ degrees of freedom when the null hypothesis is true.

If the global hypothesis test is solved applying bootstrap, the statistic for the global test is similar to that given in the expression (2.4), substituting $\hat{\boldsymbol{\tau}}$ with the bootstrap estimator of $\boldsymbol{\tau}$ and $\hat{\sum}_{\hat{\boldsymbol{\tau}}}$ with the variance-covariance matrix estimated through bootstrap.

Other alternative methods will now be proposed to solve the global hypothesis test (2.3).

2.4. Alternative methods

The method proposed in the previous Section to solve the global hypothesis test (2.3) is based on the chi-squared distribution. The following are some alternative methods to solve this hypothesis test:

Method 1. Consists of solving the $J\,(J-1)$ marginal hypothesis tests given by

$$H_0\colon\ PPV_k = PPV_l \quad \text{vs} \quad H_1\colon\ PPV_k \neq PPV_l$$

and

$$H_0\colon\ NPV_k = NPV_l \text{ vs } H_1\colon\ NPV_k \neq NPV_l$$

with $k, l = 1, ..., J$ and $k \neq l$, each one to an error rate of $\alpha/\{J\,(J-1)\}$, *i.e.* applying the Bonferroni method [17], where the statistics is

$$z = \frac{\widehat{PV_k} - \widehat{PV_l}}{\sqrt{\widehat{\mathrm{Var}}\left(\widehat{PV_k}\right) + \widehat{\mathrm{Var}}\left(\widehat{PV_l}\right) - 2\,\widehat{\mathrm{Cov}}\left(\widehat{PV_k}, \widehat{PV_l}\right)}} \longrightarrow N\,(0,1) \,, \tag{2.5}$$

and where PV is PPV or NPV respectively.

Method 2. Consists of solving the $J\,(J-1)$ marginal hypothesis tests and applying the multiple comparison method of Holm [18] to a global error rate of α.

Method 3. Consists of solving the $J\,(J-1)$ marginal hypothesis tests and applying the multiple comparison method of Hochberg [19] to a global error rate of α.

These methods are very easy to apply from the p-values calculated in the $J\,(J-1)$ marginal hypothesis tests. The Bonferroni method is a classic method of post hoc comparison, and the Holm and Hochberg methods are less conservative post hoc methods than the Bonferroni method. Furthermore all of the aforementioned methods can be applied both if the PVs and their variances-covariances are estimated through the maximum likelihood method and the delta method respectively, or if they are estimated through the bootstrap method.

3. SIMULATION EXPERIMENTS

Monte Carlo simulation experiments were carried out to study the type I error probability and the power of the global hypothesis proposed in Section 2.3 and of the alternative methods proposed in Section 2.4, when comparing the PVs of two and of three BDTs respectively, and both if the variance-covariance matrix is estimated through the delta method and if it is estimated through the bootstrap method. These simulation experiments were designed in a similar way to those carried out by Roldan-Nofuentes *et al.* [5], and consisted of the generation of 5000 random samples with multinomial distributions sized 50, 100, 200, 500, 1000, 2000 and 5000. For all of the study $\alpha = 5\%$ was set. All of the random samples were generated in such a way that in all of them it was possible to estimate the PVs and their variances-covariances. In the case of bootstrap, for each random sample 2000 samples with replacement were generated and from these $\hat{\hat{\boldsymbol{\tau}}}$ and $\hat{\sum}_{\hat{\hat{\boldsymbol{\tau}}}}$ were calculated. All of the random samples were generated from the PVs and the prevalence, without setting the values of sensitivity and specificity of each BDT in the following way:

1. As PVs we took the values $\{0.60, 0.65, ..., 0.90, 0.95\}$, which are quite common values in clinical practice, and as values of the disease prevalence we took the values $\{0.05, 0.10, ..., 0.90, 0.95\}$.
2. Once the PVs and the disease prevalence were set, the sensitivity and the specificity of each diagnostic test were calculated from equations (1.1), and then the maximum values of the dependence factors between the two $BDTs$ were obtained from the values of the sensitivity and specificity of each diagnostic test applying the model of Vacek [20] for two $BDTs$ and applying the model of Torrance-Rynard and Walter [21] for three $BDTs$. In Appendix A both models are summarized.
3. For two $BDTs$, as verification probabilities we took the values

$$\left(\lambda_{11} = 0.70,\ \lambda_{10} = \lambda_{01} = 0.40,\ \lambda_{00} = 0.10\right)$$

and

$$\left(\lambda_{11} = 0.95,\ \lambda_{10} = \lambda_{01} = 0.60,\ \lambda_{00} = 0.30\right),$$

which can be considered a scenario with low verification and a scenario with high verification respectively. For three *BDTs*, we took the values

$$\begin{aligned}\big(&\lambda_{111} = 0.70,\ \lambda_{110} = 0.40,\ \lambda_{101} = 0.40,\ \lambda_{100} = 0.25,\\ &\lambda_{011} = 0.40,\ \lambda_{010} = 0.25,\ \lambda_{001} = 0.25,\ \lambda_{000} = 0.05\big)\end{aligned}$$

and

$$\begin{aligned}\big(&\lambda_{111} = 1,\ \lambda_{110} = 0.80,\ \lambda_{101} = 0.80,\ \lambda_{100} = 0.40,\\ &\lambda_{011} = 0.80,\ \lambda_{010} = 0.40,\ \lambda_{001} = 0.40,\ \lambda_{000} = 0.20\big)\,,\end{aligned}$$

which can also be considered as scenarios with low and high verification.

4. In the case of two *BDTs*, the probabilities of the multinomial distributions were calculated from the equations of the model of Vacek [20] (Appendix A). In the case of three *BDTs*, the probabilities of the multinomial distributions were calculated from the model of Torrance-Rynard and Walter [21] (Appendix A).

3.1. Two *BDTs*

In Table 1 we show the results obtained for the type I errors probabilities and the powers when comparing the *PVs* of two *BDTs*, for different values of the *PVs* and for intermediate and high dependence factors, when the *PVs* are estimated through maximum likelihood and the variance-covariance matrix is estimated through the delta method (other tables with results from the simulation experiments can be requested from the authors). Regarding the type I error probability, the global hypothesis test has a type I error probability which, in general terms, fluctuates around the nominal error of 5% especially when $n \geqslant 500$. In some cases, especially when $n \leqslant 200$ and the verification probabilities are low and/or the dependence factors are high, the type I error probability may overwhelm the nominal error. This may be due to the fact that the samples are not large enough, and therefore some frequencies of the multinomial distribution which are equal to zero, and the variance-covariance matrix are not well represented. Regarding Methods 1, 2 and 3 (Bonferroni, Holm and Hochberg), the type I error probability of each one of them performs in a similar way to that of the global test, although it is usually somewhat lower than the nominal error (especially for $n \geqslant 2000$).

Regarding the power, in general it is necessary to have samples of between 500 and 1000 individuals (depending on the verification probabilities) so that the power of the global hypothesis test is high (higher than 80% or 90%). The power of the global hypothesis test increases when there is an increase in the verification probabilities; whereas the increase in the dependence factors does not have a clear effect on the power of the global hypothesis test (sometimes it increases and sometimes it decreases). Regarding Methods 1, 2 and 3, their respective powers perform in a similar way to that of the global test, although the power of each one of them is slightly lower than that of the global test.

Table 1: Type I errors probabilities and powers when comparing the *PVs* of two *BDTs* (M1: Method 1, M2: Method 2, M3: Method 3).

TYPE I ERRORS PROBABILITIES

$PPV_1 = PPV_2 = 0.80, \quad NPV_1 = NPV_2 = 0.90$
$p = 20\%$
$Se_1 = Se_2 = 0.5714, \quad Sp_1 = Sp_2 = 0.9643$

$\lambda_{11} = 0.70, \quad \lambda_{10} = 0.40, \quad \lambda_{01} = 0.40, \quad \lambda_{00} = 0.10$

n	$\varepsilon_1 = 0.075, \quad \varepsilon_0 = 0.01$				$\varepsilon_1 = 0.15, \quad \varepsilon_0 = 0.02$			
	Global test	**M1**	**M2**	**M3**	**Global test**	**M1**	**M2**	**M3**
50	0.007	0.007	0.007	0.007	0.003	0.003	0.003	0.003
100	0.039	0.035	0.035	0.039	0.044	0.039	0.039	0.044
200	0.070	0.069	0.069	0.069	0.068	0.067	0.067	0.068
500	0.066	0.065	0.065	0.066	0.069	0.068	0.068	0.069
1000	0.059	0.058	0.058	0.059	0.054	0.050	0.050	0.051
2000	0.056	0.055	0.055	0.055	0.049	0.039	0.039	0.042
5000	0.048	0.042	0.042	0.044	0.054	0.048	0.048	0.049

$\lambda_{11} = 0.95, \quad \lambda_{10} = 0.60, \quad \lambda_{01} = 0.60, \quad \lambda_{00} = 0.30$

n	$\varepsilon_1 = 0.075, \quad \varepsilon_0 = 0.01$				$\varepsilon_1 = 0.15, \quad \varepsilon_0 = 0.02$			
	Global test	**M1**	**M2**	**M3**	**Global test**	**M1**	**M2**	**M3**
50	0.022	0.021	0.021	0.021	0.011	0.011	0.011	0.011
100	0.062	0.058	0.058	0.062	0.067	0.066	0.066	0.067
200	0.068	0.067	0.067	0.068	0.069	0.068	0.068	0.068
500	0.055	0.056	0.056	0.057	0.058	0.057	0.057	0.057
1000	0.058	0.057	0.057	0.058	0.059	0.057	0.057	0.058
2000	0.054	0.052	0.052	0.053	0.048	0.043	0.043	0.044
5000	0.049	0.044	0.044	0.047	0.052	0.046	0.046	0.047

POWERS

$PPV_1 = 0.85, \quad PPV_2 = 0.90, \quad NPV_1 = 0.75, \quad NPV_2 = 0.80$
$p = 75\%$
$Se_1 = 0.9444, \quad Sp_1 = 0.50, \quad Se_2 = 0.9429, \quad Sp_2 = 0.6857$

$\lambda_{11} = 0.70, \quad \lambda_{10} = 0.40, \quad \lambda_{01} = 0.40, \quad \lambda_{00} = 0.10$

n	$\varepsilon_1 = 0.02, \quad \varepsilon_0 = 0.06$				$\varepsilon_1 = 0.04, \quad \varepsilon_0 = 0.12$			
	Global test	**M1**	**M2**	**M3**	**Global test**	**M1**	**M2**	**M3**
50	0.004	0.004	0.004	0.004	0.006	0.006	0.006	0.006
100	0.090	0.089	0.089	0.090	0.065	0.066	0.066	0.068
200	0.307	0.264	0.026	0.266	0.336	0.333	0.333	0.335
500	0.770	0.667	0.667	0.668	0.913	0.863	0.863	0.863
1000	0.984	0.940	0.940	0.940	0.999	0.998	0.998	0.998
2000	1	0.998	0.998	0.998	1	1	1	1
5000	1	1	1	1	1	1	1	1

$\lambda_{11} = 0.95, \quad \lambda_{10} = 0.60, \quad \lambda_{01} = 0.60, \quad \lambda_{00} = 0.30$

n	$\varepsilon_1 = 0.02, \quad \varepsilon_0 = 0.06$				$\varepsilon_1 = 0.04, \quad \varepsilon_0 = 0.12$			
	Global test	**M1**	**M2**	**M3**	**Global test**	**M1**	**M2**	**M3**
50	0.051	0.050	0.050	0.050	0.032	0.031	0.031	0.032
100	0.198	0.195	0.195	0.201	0.175	0.185	0.185	0.192
200	0.431	0.388	0.388	0.392	0.553	0.540	0.540	0.543
500	0.833	0.780	0.780	0.782	0.965	0.949	0.949	0.949
1000	0.990	0.978	0.978	0.978	1	1	1	1
2000	1	1	1	1	1	1	1	1
5000	1	1	1	1	1	1	1	1

Regarding the solution of the global test applying the bootstrap method, the results obtained are almost identical to those obtained through the method of maximum likelihood and the delta method. Therefore, in terms of the type I error probability and the power there is practically no difference between solving the global hypothesis test through the maximum likelihood method and the bootstrap method, although the bootstrap requires a greater computational effort.

3.2. Three *BDTs*

In Table 2 we show some of the results obtained for the type I error probability and the power when comparing the PVs of three $BDTs$, also for different values of the PVs and for intermediate and high dependence factors, when the PVs are estimated through maximum likelihood and the variance-covariance matrix is estimated through the delta method (other tables with results from the simulation experiments can be requested from the authors). For three $BDTs$ we have not considered sample sizes smaller than 100, since with smaller samples there are too many frequencies equal to 0 (above all when the prevalence is low and/or the verification probabilities are low) and it is not possible to calculate the estimators or the variances-covariances. In general terms, the conclusions reached are similar to those obtained for two $BDTs$, although for the global test and for methods 1, 2 and 3 it is necessary to have larger sample sizes so that the type I error probability fluctuates around the nominal error.

With regard to the power of each method, this increases with an increase in the verification probabilities, and decreases when there is an increase in the values of the dependence factors. In very general terms, when the verification probabilities are low it is necessary to have samples of between 500 and 1000 individuals so that the power of the global test is higher than 80% or 90% (depending on the values of the dependence factors), although in some situations (high dependence factors) it is necessary to have very large samples ($n \geqslant 5000$) in order to reach this power. Regarding Methods 1, 2 and 3, in general terms there is no important difference in power in relation to the global hypothesis test, especially when $n \geqslant 500$, whilst for smaller sample sizes the global test is somewhat more powerful than for the other three methods.

3.3. Conclusions

From the analysis of the results obtained in the simulation experiments one may conclude that the global hypothesis test based on the chi-squared distribution displays the performance of an asymptotic hypothesis test (from a certain sample size onwards, its type I error probability fluctuates around the nominal error).

Table 2: Type I errors probabilities and powers when comparing the *PVs* of three *BDTs* (M1: Method 1, M2: Method 2, M3: Method 3).

TYPE I ERRORS PROBABILITIES

$PPV_1 = PPV_2 = PPV_3 = 0.85, \quad NPV_1 = NPV_2 = NPV_3 = 0.80$
$p = 25\%$
$e_1 = Se_2 = Se_3 = 0.2615, \quad Sp_1 = Sp_2 = Sp_3 = 0.9846$

$\lambda_{111} = 0.70, \quad \lambda_{110} = 0.40, \quad \lambda_{101} = 0.40, \quad \lambda_{100} = 0.25$
$\lambda_{011} = 0.40, \quad \lambda_{010} = 0.25, \quad \lambda_{001} = 0.25, \quad \lambda_{000} = 0.05$

n	$\delta = 0.0075, \quad \varepsilon = 0.0001$				$\delta = 0.015, \quad \varepsilon = 0.0002$			
	Global test	**M1**	**M2**	**M3**	**Global test**	**M1**	**M2**	**M3**
100	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
200	0.005	0.003	0.003	0.003	0.005	0.004	0.004	0.004
500	0.041	0.026	0.026	0.028	0.045	0.030	0.030	0.032
1000	0.058	0.041	0.041	0.045	0.059	0.040	0.040	0.045
2000	0.059	0.043	0.043	0.050	0.059	0.058	0.058	0.059
5000	0.053	0.044	0.044	0.048	0.054	0.037	0.037	0.042

$\lambda_{111} = 1, \quad \lambda_{110} = 0.80, \quad \lambda_{101} = 0.80, \quad \lambda_{100} = 0.40$
$\lambda_{011} = 0.80, \quad \lambda_{010} = 0.40, \quad \lambda_{001} = 0.40, \quad \lambda_{000} = 0.20$

n	$\delta = 0.0075, \quad \varepsilon = 0.0001$				$\delta = 0.015, \quad \varepsilon = 0.0002$			
	Global test	**M1**	**M2**	**M3**	**Global test**	**M1**	**M2**	**M3**
100	0.007	0.005	0.005	0.005	0.006	0.004	0.004	0.005
200	0.043	0.031	0.031	0.032	0.043	0.029	0.029	0.031
500	0.050	0.037	0.038	0.041	0.056	0.039	0.039	0.041
1000	0.057	0.045	0.045	0.047	0.057	0.042	0.042	0.047
2000	0.058	0.044	0.044	0.046	0.059	0.047	0.047	0.052
5000	0.054	0.037	0.037	0.041	0.052	0.041	0.041	0.044

POWERS

$PPV_1 = 0.70, \quad PPV_2 = 0.80, \quad PPV_3 = 0.90, \quad NPV_1 = 0.75, \quad NPV_2 = 0.85, \quad NPV_3 = 0.95$
$p = 60\%$
$Se_1 = 0.9074, \quad Sp_1 = 0.4167, \quad Se_2 = 0.9231, \quad Sp_2 = 0.6538, \quad Se_3 = 0.9706, \quad Sp_3 = 0.8382$

$\lambda_{111} = 0.70, \quad \lambda_{110} = 0.40, \quad \lambda_{101} = 0.40, \quad \lambda_{100} = 0.25$
$\lambda_{011} = 0.40, \quad \lambda_{010} = 0.25, \quad \lambda_{001} = 0.25, \quad \lambda_{000} = 0.05$

n	$\delta = 0.001, \quad \varepsilon = 0.01$				$\delta = 0.002, \quad \varepsilon = 0.02$			
	Global test	**M1**	**M2**	**M3**	**Global test**	**M1**	**M2**	**M3**
100	0.042	0.041	0.041	0.041	0.034	0.033	0.033	0.034
200	0.261	0.260	0.260	0.260	0.219	0.218	0.218	0.218
500	0.823	0.823	0.823	0.823	0.741	0.740	0.740	0.740
1000	0.986	0.986	0.986	0.986	0.950	0.950	0.950	0.950
2000	1	1	1	1	1	1	1	1
5000	1	1	1	1	1	1	1	1

$\lambda_{111} = 1, \quad \lambda_{110} = 0.80, \quad \lambda_{101} = 0.80, \quad \lambda_{100} = 0.40$
$\lambda_{011} = 0.80, \quad \lambda_{010} = 0.40, \quad \lambda_{001} = 0.40, \quad \lambda_{000} = 0.20$

n	$\delta = 0.001, \quad \varepsilon = 0.01$				$\delta = 0.002, \quad \varepsilon = 0.02$			
	Global test	**M1**	**M2**	**M3**	**Global test**	**M1**	**M2**	**M3**
100	0.316	0.315	0.315	0.315	0.276	0.275	0.275	0.275
200	0.724	0.724	0.724	0.724	0.652	0.651	0.651	0.651
500	0.971	0.970	0.970	0.971	0.910	0.909	0.909	0.910
1000	1	1	1	1	0.995	0.995	0.995	0.995
2000	1	1	1	1	1	1	1	1
5000	1	1	1	1	1	1	1	1

In general terms, its type I error probability fluctuates around the nominal error (especially for $n \geqslant 500$) and it is necessary to have large samples ($n \geqslant 500$) so that the power is greater than 80%. From the results obtained in the simulation experiments carried out, the global hypothesis test

$$\begin{aligned} H_0 &: \; PPV_1 = PPV_2 = \cdots = PPV_J \quad \text{and} \quad NPV_1 = NPV_2 = \cdots = NPV_J \;, \\ H_1 &: \; \text{at least one equality is not true} \;, \end{aligned}$$

can be solved through the following procedure:

1. Solving the global hypothesis test based on the chi-squared distribution to a global error rate of α using the statistics given by equations (2.4) or bootstrap method.
2. If the global hypothesis test is not significant, then one cannot reject the homogeneity of the J *PPVs* and of the J *NPVs*. If the global hypothesis test is significant to an error rate of α, in order to investigate the causes of the significance the following marginal hypothesis tests are solved

   $$H_0\colon \; PPV_i = PPV_j \quad \text{vs} \quad H_1\colon \; PPV_i \neq PPV_j$$
   and
   $$H_0\colon \; NPV_i = NPV_j \quad \text{vs} \quad H_1\colon \; NPV_i \neq NPV_j$$

   using the statistics given by equation (2.5), and applying some of the methods of multiple comparison used (Bonferroni, Holm or Hochberg) to an error rate of α.

4. EXAMPLE

The results obtained in Section 2 and the procedure given in Section 3.3 were applied to the diagnosis of coronary stenosis. Coronary stenosis is a disease that consists of the obstruction of the coronary artery and its diagnosis can be made through a dobutamine echocardiogram, a stress echocardiogram or a *CT* scan, and as the gold standard a coronary angiography is used. Coronary angiography may cause different reactions in patients (thrombosis, heart attacks, infections, even death) and therefore not all patients are verified with the gold standard. In Table 3 (Study of coronary stenosis), we show the results obtained by applying the dobutamine echocardiogram (variable T_1), the stress echocardiogram (variable T_2) and the CT scan (variable T_2) to a sample of 2455 males over 45 years of age and by only applying the coronary angiography (variable D) to a subset of these individuals. This study was carried out in two phases: firstly, the three *BDTs* were applied to all of the individuals in the sample, and secondly the gold standard was applied to a subset of these individuals depending on only the results of the three diagnostic tests. This data are part of a study carried out at the University Hospital in Granada (Spain). In this example, one can assume that the missing data mechanism is ignorable, and therefore the results from Section 3 can be applied. The values of the estimators of the PVs are $\widehat{PPV_1} = 0.742$, $\widehat{PPV_2} = 0.622$, $\widehat{PPV_3} = 0.805$, $\widehat{NPV_1} = 0.933$, $\widehat{NPV_2} = 0.850$, $\widehat{NPV_3} = 0.952$, and applying the delta method, the estimated asymptotic variance-covariance matrix is

$$\hat{\sum}_{\hat{\tau}} = \begin{pmatrix} 0.000234 & 0.000108 & 0.000086 & 0 & -0.000063 & -0.000038 \\ 0.000108 & 0.000258 & 0.000106 & -0.000035 & 0 & -0.000025 \\ 0.000086 & 0.000106 & 0.0000239 & -0.000059 & -0.000069 & 0 \\ 0 & -0.000034 & -0.000059 & 0.000114 & 0.000080 & 0.000045 \\ 0.000063 & 0 & 0.000069 & 0.000080 & 0.000169 & 0.000064 \\ 0.000038 & 0.000025 & 0 & 0.000045 & 0.000064 & 0.000085 \end{pmatrix}.$$

Applying equation (2.4) it holds that $Q^2 = 145.103$ ($p - value = 0$), and therefore we reject the equality of the three $PPVs$ and of the three $NPVs$. In order to investigate the causes of the significance, the marginal hypothesis tests ($H_0\colon PPV_i = PPV_j$ and $H_0\colon NPV_i = NPV_j$) are solved. In Table 3 (Marginal hypothesis tests), we show the results obtained for each one of the six hypothesis tests that compare the PVs. Applying the Bonferroni method, the Holm method or the Hochberg method, it holds that the three $PPVs$ are different, and that the PPV of the CT scan is the largest, followed by that of the dobutamine echocardiogram and, finally, that of the stress echocardiogram.

Table 3: Data from the study of coronary stenosis and marginal hypothesis tests.

Study of coronary stenosis									
	$T_1 = 1$				$T_1 = 0$				
	$T_2 = 1$		$T_2 = 0$		$T_2 = 1$		$T_2 = 0$		
	$T_3 = 1$	$T_3 = 0$	$T_3 = 1$	$T_3 = 0$	$T_3 = 1$	$T_3 = 0$	$T_3 = 1$	$T_3 = 0$	Total
V = 1									
D = 1	457	30	84	5	34	0	7	1	618
D = 0	41	23	5	61	16	86	32	95	359
V = 0	92	31	85	120	42	195	88	825	1478
Total	590	84	174	186	92	281	127	921	2455

Marginal hypothesis tests		
Hypothesis test	z	Two sided p-value
$H_0\colon PPV_1 = PPV_2$ vs $H_1\colon PPV_1 \neq PPV_2$	3.61	0.003
$H_0\colon PPV_1 = PPV_3$ vs $H_1\colon PPV_1 \neq PPV_3$	7.20	6.06×10^{-13}
$H_0\colon PPV_2 = PPV_3$ vs $H_1\colon PPV_2 \neq PPV_3$	10.79	0
$H_0\colon NPV_1 = NPV_2$ vs $H_1\colon NPV_1 \neq NPV_2$	7.46	8.37×10^{-14}
$H_0\colon NPV_1 = NPV_3$ vs $H_1\colon NPV_1 \neq NPV_3$	1.76	0.078
$H_0\colon NPV_2 = NPV_3$ vs $H_1\colon NPV_2 \neq NPV_3$	8.99	0

Regarding the $NPVs$, no significant differences were found between the $NPVs$ of the dobutamine echocardiogram and of the CT scan, whilst the NPV of the dobutamine echocardiogram is significantly lower than the $NPVs$ of the other two $BDTs$.

5. DISCUSSION

Different studies have examined the problem of the comparison of the PVs of two or more $BDTs$ when the diagnostic tests and the gold standard are applied to all of the individuals in a random sample. These models cannot be applied when a subset of individuals in the random sample have not had their disease status verified through the application of the gold standard, since the results obtained may be affected by the verification bias. In this article, we have studied a global hypothesis test to simultaneously compare the PVs of two or more $BDTs$ when for a subset of individuals in the sample the disease status (either present or absent) is unknown. The global hypothesis test is based on the chi-squared distribution, and can be solved through the method of maximum likelihood and the delta method (equation (2.4) or through the bootstrap method, although the latter requires a greater computational effort. In terms of the type I error probability, both methods lead to very similar results, and the type I error probability fluctuates around the nominal error especially for $n \geqslant 500$. Other alternative methods to solve the global hypothesis test have been studied. The method based on the marginal comparisons of the $PPVs$ ($NPVs$) to an error rate of $\alpha = 5\%$ leads to a type I error probability that clearly overwhelms the nominal error, and therefore this method may give rise to erroneous results. The methods based on marginal comparisons applying the corrections of Bonferroni, Holm and Hochberg respectively give rise to a type I error probability that fluctuates around the nominal error especially for $n \geqslant 500$. In terms of power, the global hypothesis test based on the chi-squared distribution (equation (2.4) or bootstrap method) is a little more powerful than the methods based on the corrections of Bonferroni, Holm and Hochberg respectively. Therefore, from the results of the simulation experiments carried out, the following method is proposed to compare the PVs of J $BDTs$ in the presence of ignorable missing data: 1) Apply the global hypothesis test based on the chi-squared distribution to an error rate of α (equations (2.4) or bootstrap method); 2) If the global hypothesis test is significant to an error rate of α, investigate the causes of the significance solving the marginal hypothesis tests $H_0\colon PPV_i = PPV_j$ and $H_0\colon NPV_i = NPV_j$ along with a method of multiple comparisons (Bonferroni, Holm or Hochberg). This procedure is similar to that used in an analysis of variance. Firstly, the global test is solved and then a method of multiple comparisons is applied.

An alternative method to that proposed in Section 2 consists of solving the global test applying the Wilks method. Similar simulation experiments to those described in Section have demonstrated that the type I error probability and the power of this method are very similar to those obtained with the Wald method (equation (2.4)).

If all of the individuals are verified with the gold standard, and therefore all of the frequencies $u_{i_1,\ldots,i_J}$ are equal to 0, the method proposed by Roldán-Nofuentes *et al.* [5] is a particular case of the scenario analyzed in this study. Therefore, the simultaneous comparison of the PVs of two (or more) $BDTs$ in paired designs is a particular case of the scenario analyzed in this article.

APPENDIX A

In the case of two *BDTs*, the probabilities of the multinomial distribution were calculated applying the model of conditional dependence of Vacek [20], and their expressions are

$$\begin{aligned}
\phi_{ij} &= \lambda_{ij} p \left\{ Se_1^i (1-Se_1)^{1-i} Se_2^j (1-Se_2)^{1-j} + \delta_{ij}\varepsilon_1 \right\} , \\
\varphi_{ij} &= \lambda_{ij} (1-p) \left\{ Sp_1^{1-i} (1-Sp_1)^i Sp_2^{1-j} (1-Sp_2)^j + \delta_{ij}\varepsilon_0 \right\} , \\
\gamma_{ij} &= (1-\lambda_{ij})\, p \left\{ Se_1^i (1-Se_1)^{1-i} Se_2^j (1-Se_2)^{1-j} + \delta_{ij}\varepsilon_1 \right\} \\
&\quad + (1-\lambda_{ij})\,(1-p) \left\{ Sp_1^{1-i} (1-Sp_1)^i Sp_2^{1-j} (1-Sp_2)^j + \delta_{ij}\varepsilon_0 \right\} ,
\end{aligned}$$

where $\delta_{ij} = 1$ when $i = j$ and $\delta_{ij} = -1$, and ε_i is the dependence factor (covariance) between the two *BDTs* when $D = i$. In clinical practice, the two *BDTs* are usually conditionally dependent on the disease, and it is verified [20] that $0 < \varepsilon_1 < Se_1 (1 - Se_2)$ when $Se_2 > Se_1$ and $0 < \varepsilon_1 < Se_2 (1 - Se_1)$ when $Se_1 > Se_2$, and in the same way, $0 < \varepsilon_0 < Sp_1 (1 - Sp_2)$ when $Sp_2 > Sp_1$ and $0 < \varepsilon_0 < Sp_2 (1 - Sp_1)$ when $Sp_1 > Sp_2$. If the two *BDTs* are conditionally independent on the disease then $\varepsilon_1 = \varepsilon_0 = 0$.

In the case of three *BDTs*, the probabilities of the multinomial distributions were calculated applying the model of Torrance-Rynard and Walter [21]:

$$\begin{aligned}
&P\Big(V=1,\, D=1,\, T_1=i_1,\, T_2=i_2,\, T_3=i_3\Big) = \\
&\qquad = p\,\lambda_{i_1 i_2 i_3} \left\{ \prod_{j=1}^{3} Se_j^{i_j} (1-Se_j)^{1-i_j} + \sum_{j,k,j<k}^{3} (-1)^{|i_j - i_k|} \delta_{jk} \right\} ,
\end{aligned}$$

$$\begin{aligned}
&P\Big(V=1,\, D=0,\, T_1=i_1,\, T_2=i_2,\, T_3=i_3\Big) = \\
&\qquad = (1-p)\,\lambda_{i_1 i_2 i_3} \left\{ \prod_{j=1}^{3} {Sp_j}^{1-i_j} (1-Sp_j)^{i_j} + \sum_{j,k,j<k}^{3} (-1)^{|i_j - i_k|} \varepsilon_{jk} \right\}
\end{aligned}$$

and

$$\begin{aligned}
&P\Big(V=0,\, T_1=i_1,\, T_2=i_2,\, T_3=i_3\Big) = \\
&\quad = p\,(1-\lambda_{i_1 i_2 i_3}) \left\{ \prod_{j=1}^{3} {Se_j}^{i_j} (1-Se_j)^{1-i_j} + \sum_{j,k,j<k}^{3} (-1)^{|i_j - i_k|} \delta_{jk} \right\} \\
&\qquad + (1-p)\,(1-\lambda_{i_1 i_2 i_3}) \left\{ \prod_{j=1}^{3} {Sp_j}^{1-i_j} (1-Sp_j)^{i_j} + \sum_{j,k,j<k}^{3} (-1)^{|i_j - i_k|} \varepsilon_{jk} \right\} ,
\end{aligned}$$

with $i_j = 0, 1,\ i_k = 0, 1$ and $j, k = 1, 2, 3,$ and where δ_{jk} (ε_{jk}) is the factor of dependence between the j-th *BDT* and k-th *BDT* when $D = 1$ $(D = 0)$.

The factors of dependence δ_{jk} and ε_{jk} verify restrictions that depend on the values of sensitivity and specificity of the three *BDTs*. In order to simplify the simulation experiments, it has been considered that $\delta_{ij} = \delta$ and $\varepsilon_{ij} = \varepsilon$, so that the factors of dependence verify the following restrictions:

$$\delta \le \mathrm{Min}\Big\{(1-Se_1)\,(1-Se_2)\,Se_3,\ (1-Se_1)\,Se_2\,(1-Se_3)\,,\ Se_1\,(1-Se_2)\,(1-Se_3)\Big\}$$

and

$$\varepsilon \le \mathrm{Min}\Big\{(1-Sp_1)\,(1-Sp_2)\,Sp_3,\ (1-Sp_1)\,Sp_2\,(1-Sp_3)\,,\ Sp_1\,(1-Sp_2)\,(1-Sp_3)\Big\}.$$

In clinical practice, factors δ_{jk} and/or ε_{jk} are greater than zero, so that the *BDTs* are conditionally dependent on the disease status. When $\delta_{jk} = \varepsilon_{jk} = 0$ the three BDTs are conditionally independent on the disease status.

ACKNOWLEDGMENTS

This research was supported by the Spanish Ministry of Economy, Grant Number MTM2012-35591. We thank the two referees, the Associate Editor and the Editor (M. Ivette Gomes) of REVSTAT for their helpful comments that improved the quality of the paper.

REFERENCES

[1] Bennet, B. M. (1972). On comparison of sensitivity, specificity and predictive value of a number of diagnostic procedures, *Biometrics*, **28**, 793–800.

[2] Bennet, B. M. (1985). On tests for equality of predictive values for t diagnostic procedures, *Statistics in Medicine*, **4**, 525–539.

[3] Leisenring, W.; Alonzo, T. and Pepe, M. S. (2000). Comparisons of predictive values of binary medical diagnostic tests for paired designs, *Biometrics*, **56**, 354–351.

[4] Wnag, W.; Davis, C. S. and Soong, S. J. (2006). Comparison of predictive values of two diagnostic tests from the same sample of subjects using weighted least squares, *Statistics in Medicine*, **25**, 2215–2229.

[5] Roldán Nofuentes, J. A.; Luna del Castillo, J. D. and Montero Alonso, M. A. (2012). Global hypothesis test to simultaneously compare the predictive values of two binary diagnostic tests, *Computational Statistics and Data Analysis, Special issue "Computational Statistics for Clinical Research"*, **56**, 1161–1173.

[6] Poleto, F. Z.; Singer, J. M. and Paulino, C. D. (2011). Comparing diagnostic tests with missing data, *Journal of Applied Statistics*, **38**, 1207–1222.

[7] Begg, C. B. and Greenes, R. A. (1983). Assessment of diagnostic tests when disease verification is subject to selection bias, *Biometrics*, **39**, 207–215.

[8] Zhou, X. H. (1993). Maximum likelihood estimators of sensitivity and specificity corrected for verification bias, *Communication in Statistics Theory and Methods*, **22**, 3177–3198.

[9] Roldán Nofuentes, J. A. and Luna del Castillo, J. D. (2007). The effect of verification bias in the naive estimators of accuracy of a binary diagnostic test, *Communications in Statistics — Simulation and Computation*, **36**, 959–972.

[10] Roldán Nofuentes, J. A. and Luna del Castillo, J. D. (2008a). EM algorithm for comparing two binary diagnostic tests when not all the patients are verified, *Journal of Statistical Computation and Simulation*, **78**, 19–35.

[11] Roldán Nofuentes, J. A. and Luna del Castillo, J. D. (2008b). The effect of verification bias on the comparison of predictive values of two binary diagnostic tests, *Journal of Statistical Planning and Inference*, **138**, 959–963.

[12] Rubin, D. B. (1976). Inference and missing data, *Biometrika*, **4**, 73–89.

[13] Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*, Chapman and Hall/CRC, USA.

[14] Bishop, Y. M.; Fienberg, S. E. and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*, The MIT Press.

[15] Agresti, A. (2002). *Categorical Data Analysis*, John Wiley and Sons, New York.

[16] Efron, B. and Tibshirsni, R. J. (1993). *An Introduction to the Bootstrap*, Chapman and Hall, New York.

[17] Bonferroni, C. E. (1936). Teoria statistica delle classi e calcolo delle probabilitá, *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, **8**, 3–62.

[18] Holm, S. (1979). A simple sequential rejective multiple testing procedure, *Scandinavian Journal of Statistics*, **6**, 65–70.

[19] Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance, *Biometrika*, **75**, 800–802.

[20] Vacek, P. (1985). The effect of conditional dependence on the evaluation of diagnostic tests, *Biometrics*, **23**, 959–968.

[21] Torrance-Rynard, V. L. and Walter, S. D. (1997). Effects of dependent errors in the assessment of diagnostic test performance, *Statistics in Medicine*, **16**, 2157–2175.

REVSTAT – Statistical Journal
Volume 15, Number 1, January 2017, 65–88

THE BETA GENERALIZED INVERTED EXPONENTIAL DISTRIBUTION WITH REAL DATA APPLICATIONS

Authors: R.A. Bakoban
– Department of Statistics, Faculty of Science AL-Faisaliah Campus, King Abdulaziz University, Jeddah, Saudi Arabia
rbakoban@kau.edu.sa

Hanaa H. Abu-Zinadah
– Department of Statistics, Faculty of Science, AL-Faisaliah Campus, King Abdulaziz University, P.O. Box 32691, Jeddah 21438, Saudi Arabia
habuzinadah@kau.edu.sa

Received: July 2014 Revised: April 2015 Accepted: May 2015

Abstract:

- The four-parameter beta generalized inverted exponential distribution is considered in this article. Various properties of the model with graphs of the density function are investigated. Moreover, the maximum likelihood method of estimation is used for estimating the parameters of the model under complete samples. An asymptotic Fisher information matrix of the estimators is found. Additionally, confidence interval estimates of the parameters are obtained. The performances of findings of the article are shown by demonstrating various numerical illustrations through Monte Carlo simulation studies. Finally, applications on real data-sets are provided.

Key-Words:

- *beta generalized inverted exponential distribution; Fisher information matrix; goodness-of-fit test; maximum likelihood estimator; Monte Carlo simulation.*

AMS Subject Classification:

- 62-07, 62E20, 62F10, 62F25, 62N02.

1. INTRODUCTION

The generalized inverted exponential distribution (GIED) was introduced first by Abouammoh and Alshingiti (2009). It is a generalized form of the inverted exponential distribution (IED). IED has been studied by Keller and Kamath (1982) and Duran and Lewis (1989). GIED has good statistical and reliability properties. It fits various shapes of failure rates.

The probability density function (pdf) of a two-parameter GIED is given by

$$f(x) = \left(\frac{\alpha\lambda}{x^2}\right) \exp\left(\frac{-\lambda}{x}\right) \left[1 - \exp\left(\frac{-\lambda}{x}\right)\right]^{\alpha-1}, \qquad x > 0, \quad \alpha, \lambda > 0\,, \tag{1.1}$$

and the cumulative distribution function (cdf) is given by

$$F(x) = 1 - \left[1 - \exp\left(\frac{-\lambda}{x}\right)\right]^{\alpha}, \qquad x > 0, \quad \alpha, \lambda > 0\,. \tag{1.2}$$

In the last few years, new classes of distributions have been found by extending certain distributions such that these new classes will have more applications in reliability, biology and other fields.

Let $G(t)$ be a cdf of a random variable T, such that

$$F(t) = \frac{1}{B(a,b)} \int_0^{G(t)} \varpi^{a-1} (1-\varpi)^{b-1} d\varpi\,, \tag{1.3}$$

where $a > 0$, $b > 0$, and $B(a,b) = \int_0^1 \varpi^{a-1}(1-\varpi)^{b-1} d\varpi$ is the beta function. The skewness of the distribution is controlled by the two parameters a and b. The cdf $G(t)$ could be any arbitrary distribution, and, consequently, F is named the beta G distribution. The previous formula in (1.3) was defined by Eugene *et al.* (2002) as a class of generalized distributions.

The beta normal distribution (BND) was introduced by Eugene *et al.* (2002). They used the cdf $G(t)$ of the normal distribution in (1.3) and derived some moments of the distribution. Expanding on this work, Gupta and Nadarajah (2004) established more general moments of BND. Based on the cdf $G(t)$ of the Gumbel distribution, Nadarajah and Kotz (2004) presented the beta Gumbel distribution and provided closed form expressions for the moments and the asymptotic distribution of the extreme order statistics and obtained the maximum likelihood estimators (MLE) of the parameters. Further, by using the cdf $G(t)$ of the exponential distribution, Nadarajah and Kotz (2005) considered the beta exponential distribution. They studied the first four cumulants, the moment generating function, and the extreme order statistics and found the MLE. Furthermore, Lee *et al.* (2007) considered the beta Weibull distribution and studied applications based on censored data.

Recently, Barreto-Souza *et al.* (2010) proposed the beta generalized exponential distribution by taking G(t) in (1.3) to be the cdf of the exponentiated exponential distribution and discussed the MLE of its parameters. Additionally, Nassar and Nada (2011) presented several properties of the beta generalized Pareto distribution. They estimated the distribution's parameters using the MLE. An application on actual tax revenue data was investigated. Paranaiba *et al.* (2011) discussed the beta Burr XII distribution. Mahmoudi (2011) presented the beta generalized Pareto distribution. Cordeiro and Lemonte (2011) investigated the beta Laplace distribution. Zea *et al.* (2012) studied statistical properties and inference of the beta exponentiated Pareto distribution (BEPD). They provided an application of the BEPD to remission times of bladder cancer. Leão *et al.* (2013) studied the beta inverse Rayleigh distribution. They provided various properties, including the quantile function, moments, mean deviations, Bonferroni and Lorenz curves, Rényi and Shannon entropies and order statistics, as well as the MLE. Baharith *et al.* (2014) discussed properties, the MLE and the Fisher information matrix for the beta generalized inverse Weibull distribution.

In this paper, a new beta distribution is introduced by taking $G(\cdot)$ to be the GIED, and we refer to it as the beta generalized inverted exponential distribution (BGIED). In Section 2, the BGIED is defined. Statistical properties of the model are derived in Section 3. Maximum likelihood estimators of the parameters are derived in Section 4. In Section 5, the asymptotic Fisher information matrix is investigated. Additionally, interval estimates of the parameters are found using the maximum likelihood method in Section 6. Section 7 explains the simulation studies that illustrate the theoretical results. Finally, Section 8 provides applications to real data-sets. Various conclusions are addressed in Section 9.

2. BETA GENERALIZED INVERTED EXPONENTIAL DISTRIBUTION

In this section, we introduce the four-parameter beta generalized inverted exponential distribution (BGIED) by assuming $G(x)$ to be the cdf of the generalized inverted exponential distribution (GIED). Substituting (1.2), the cdf of GIED, into (1.3), the cdf of the BGIED is obtained in the following form

$$F(x) = \frac{1}{B(a,b)} \int_0^{1-\left[1-\exp\left(\frac{-\lambda}{x}\right)\right]^\alpha} \varpi^{a-1}(1-\varpi)^{b-1} d\varpi \,, \qquad x>0, \quad a,b,\alpha \text{ and } \lambda>0 \,. \tag{2.1}$$

The pdf of the BGIED takes the form

$$f(x) = \frac{\alpha\lambda\exp\left(\frac{-\lambda}{x}\right)}{x^2 B(a,b)} \left(1-\left[1-\exp\left(\frac{-\lambda}{x}\right)\right]^\alpha\right)^{a-1} \left[1-\exp\left(\frac{-\lambda}{x}\right)\right]^{\alpha b-1}, \qquad x>0, \quad a,b,\alpha \text{ and } \lambda>0 \,. \tag{2.2}$$

For a positive real value $a > 0$, (2.2) can be rewritten as an infinite power series in the form

$$f(x) = \frac{\alpha\lambda \exp\left(\frac{-\lambda}{x}\right)}{x^2 B(a,b)} \sum_{k=0}^{\infty} \frac{(-1)^k \Gamma(a)}{k!\Gamma(a-k)} \left[1 - \exp\left(\frac{-\lambda}{x}\right)\right]^{\alpha(b+k)-1}, \tag{2.3}$$

$$x > 0, \quad a, b, \alpha, \quad \text{and } \lambda > 0 .$$

From (2.3), the corresponding cdf can be written as follows

$$F(x) = \frac{1}{B(a,b)} \sum_{k=0}^{\infty} \frac{(-1)^{k+1}}{a(b+k)B(a-k,k+1)} \left[1 - \exp\left(\frac{-\lambda}{x}\right)\right]^{\alpha(b+k)}, \tag{2.4}$$

$$x > 0, \quad a, b, \alpha \text{ and } \lambda > 0 .$$

The GIED is a special case of (2.2) when $a = b = 1$. Therefore, we can assume all of the properties of the GIED that were investigated by Abouammoh and Alshingiti (2009) still hold. Additionally, when $\alpha = 1$ in (2.2), the BIED is obtained, which is related to the BGIWD when the shape parameters are equal to one and has been discussed by Baharith *et al.* (2014).

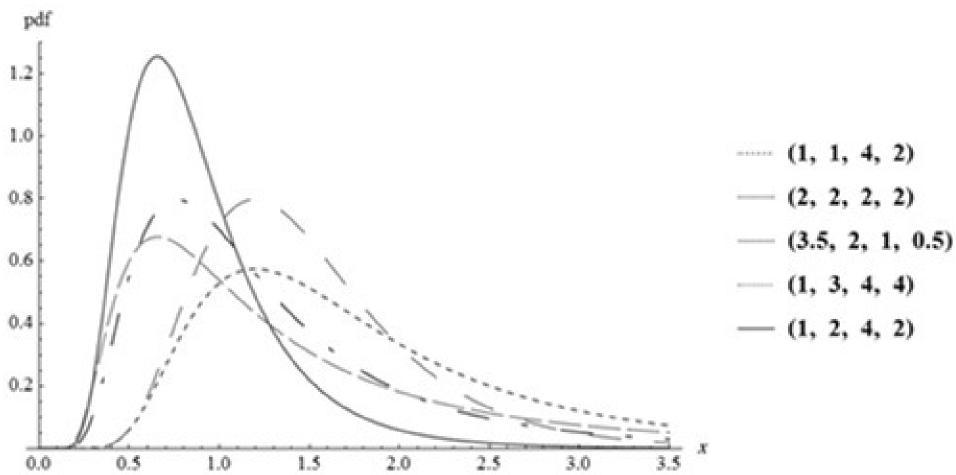


Figure 1: The pdf curves of the BGIED with (a, b, α, λ).

3. STATISTICAL PROPERTIES

3.1. The reliability and hazard functions

The reliability function is the probability of no failure occurring before time t. Alternately, the hazard function is the instantaneous rate of failure at a given time. These two functions are very important properties of a lifetime distribution.

The reliability function of the BGIED is given by

$$R(x) = 1 - \frac{1}{B(a,b)} \sum_{k=0}^{\infty} \frac{(-1)^{k+1}}{a(b+k)B(a-k,k+1)} \left[1 - \exp\left(\frac{-\lambda}{x}\right)\right]^{\alpha(b+k)}, \tag{3.1}$$

$$x > 0, \quad a, b, \alpha \text{ and } \lambda > 0\,,$$

and the corresponding hazard function of the BGIED can be written as

$$h(x) = \frac{\frac{\alpha\lambda \exp\left(\frac{-\lambda}{x}\right)}{x^2 B(a,b)} \sum_{k=0}^{\infty} \frac{(-1)^k \Gamma(a)}{k!\Gamma(a-k)} \left[1 - \exp\left(\frac{-\lambda}{x}\right)\right]^{\alpha(b+k)-1}}{1 - \frac{1}{B(a,b)} \sum_{k=0}^{\infty} \frac{(-1)^{k+1}}{a(b+k)B(a-k,k+1)} \left[1 - \exp\left(\frac{-\lambda}{x}\right)\right]^{\alpha(b+k)}}, \tag{3.2}$$

$$x > 0, \quad a, b, \alpha \text{ and } \lambda > 0\,.$$

Figure 2 shows different choices for the parameters of the BGIED. Additionally, it is shown from Figure 3 that the hazard function of the BGIED has an upside down bathtub shape. As is shown, the hazard function increases and then decreases.

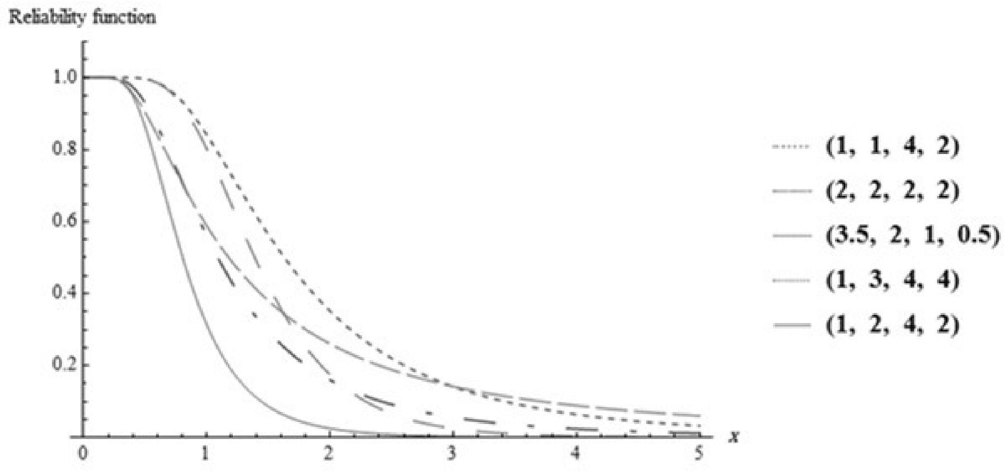


Figure 2: The reliability curves of the BGIED with (a, b, α, λ).

The upside down bathtub hazard function indicates that the risk of failing decreases as soon as the item has passed a specific time, during which it may have experienced some type of stress. Thus, the BGIED shows good statistical behavior based on these two functions.

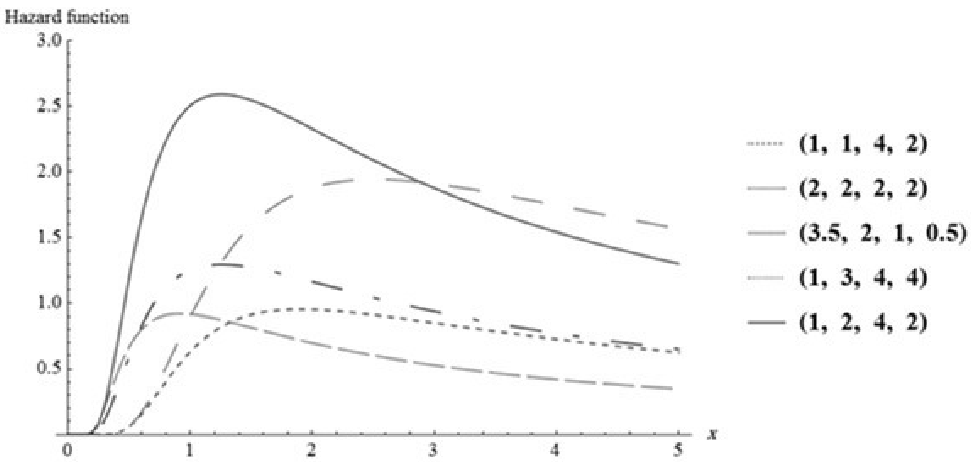


Figure 3: The hazard curves of the BGIED with (a, b, α, λ).

3.2. Moments and various measures

The r^{th} moment about the origin, $\mu'_r = E(X^r)$ of a BGIED with pdf (2.2) in the non-closed form is

$$\mu'_r = \int_0^\infty x^r \frac{\alpha\lambda \exp\left(\frac{-\lambda}{x}\right)}{x^2 B(a,b)} \left(1 - \left[1 - \exp\left(\frac{-\lambda}{x}\right)\right]^\alpha\right)^{a-1} \left[1 - \exp\left(\frac{-\lambda}{x}\right)\right]^{\alpha b - 1} dx\,, \qquad r = 1, 2, ...$$

that is, for $k \geq r$, μ'_r takes the closed form

$$\text{(3.3)} \qquad \mu'_r = \frac{\lambda^r}{B(a,b)} \sum_{k=0}^\infty \sum_{j=0}^\infty \frac{(-1)^{k+j}(j+1)^{r-1}}{a(b+k)B(a-k,k+1)B(j+1,\alpha(b+k)-j)} \times \left\{\sum_{i=0}^\infty \frac{(-1)^i}{i!(i-r+1)} + E_r(1)\right\},$$

where $B(a,b)$ is the beta function, and $E_n(z)$ is called the exponential integral function (Abramowitz and Stegun (1972)), which is defined as

$$\text{(3.4)} \qquad E_n(z) = \int_1^\infty \frac{\exp(-zt)}{t^n}\, dt\,.$$

Substituting $r = 1$ in (3.3), we obtain the mean of the BGIED as follows

$$\text{(3.5)} \qquad \mu = \frac{\lambda}{B(a,b)} \sum_{k=1}^\infty \sum_{j=0}^\infty \frac{(-1)^{k+j}}{a(b+k)B(a-k,k+1)B(j+1,\alpha(b+k)-j)} \times \left\{\sum_{i=1}^\infty \frac{(-1)^i}{i^2(i-1)!} + E_1(1)\right\},$$

where $E_1(1) = 0.577216$ is Euler's constant.

Additionally, the variance of the BGIED can be found from

$$(3.6)\quad \mathrm{Var}(x) = \frac{\lambda^2}{B(a,b)} \sum_{k=2}^{\infty}\sum_{j=0}^{\infty} \frac{(-1)^{(k+j)}(j+1)}{a(b+k)B(a-k,k+1)B(j+1,\alpha(b+k)-j)} \times \left\{\sum_{i=0}^{\infty} \frac{(-1)^i}{i!(i-1)} + E_2(1)\right\} - \mu^2 .$$

3.3. Quantile function and various related measures

The quantile function of the BGIED corresponding to (2.2) is

$$(3.7)\qquad q(u) = -\lambda \Big/ \log\left\{1 - \left[1 - I_u^{-1}(a,b)\right]^{\frac{1}{\alpha}}\right\}, \qquad 0 < u < 1 ,$$

where $I_u^{-1}(a,b)$ is the inverse of the incomplete beta function with parameters a and b, such that

$$I_u(a,b) = \frac{1}{B(a,b)} \int_0^u \varpi^{a-1}(1-\varpi)^{b-1} d\varpi ,$$

The above form of $q(u)$ allows us to derive the following forms of statistical measures for the BGIED:

1. The first quartile $Q1$, the second quartile $Q2$ (median), and the third quartile $Q3$ of the BGIED correspond to the values $u = 0.25$, 0.50, and 0.75, respectively

2. The median (m), also, can be found using (2.4) such that $\left|1-\exp\left(\frac{-\lambda}{m}\right)\right| < 1$, for $a = 1$, and then

$$(3.8)\qquad m = \frac{-\lambda}{\log\left[1 - (-0.5)^{\frac{1}{\alpha b}}\right]} .$$

3. The skewness and kurtosis can be calculated by using the following relations, respectively:

Bowley's skewness is based on quartiles; Kenney and Keeping (1962) calculated it as follows

$$(3.9)\qquad \upsilon_3 = \frac{Q3 - 2\,Q2 + Q1}{Q3 - Q1} ,$$

Moors' kurtosis (Moors (1988)) is based on octiles via the form

$$(3.10)\qquad \upsilon_4 = \frac{q(7/8) - q(5/8) - q(3/8) + q(1/8)}{q(6/8) - q(2/8)} ,$$

where $q(\cdot)$ represents the quantile function defined in (3.7).

When $a = b = 1$ in (2.3), (3.3) and (3.7) give the moments and the quantile of GIED, and, when $a = b = \alpha = 1$ in (2.3), (3.3) and (3.7) give the moments and the quantile of IED. Therefore, all measures above are satisfied for GIED when $a = b = 1$, and for IED when $a = b = \alpha = 1$.

3.4. The mean deviation

Let X be a BGIED random variable with mean $\mu = E(X)$ and median m. In this subsection, the mean deviation from the mean and the mean deviation from the median are derived.

3.4.1. The mean deviation from the mean can be found from the following theorem:

Theorem 1. *The mean deviation from the mean of the BGIED is in the form*

$$\begin{aligned} E(|X-\mu|) &= \frac{2}{B(a,b)} \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \frac{(-1)^{k+1+j}}{a(b+k)B(a-k,k+1)\left[\alpha(b+k)+1\right]} \\ &\quad \times \frac{\mu \exp\left(-j\lambda/\mu\right) - j\lambda\Gamma\left(0, j\lambda/\mu\right)}{B\left[j+1, \alpha(b+k)-j+1\right]} \,, \end{aligned}$$

where $\Gamma\left(a,z\right) = \int_z^{\infty} t^{a-1} \exp\left(-t\right) dt$.

Proof: The mean deviation from the mean can be defined as

$$\begin{aligned} \mathrm{E}(|X-\mu|) &= \int_0^{\infty} |X-\mu|\, f(x)\ dx \\ &= 2\int_0^{\mu} (X-\mu)\, f(x)\ dx \\ &= 2\mu\, F(\mu) - 2I(\mu)\ , \end{aligned}$$

where $I(z) = \int_0^z t\, dG(t)$, and $d\left[t.dG(t)\right] = G(t)\, dt + t\, dG(t)$.

Therefore, $\mathrm{E}(|X-\mu|) = 2\int_0^{\mu} F(x)\, dx$.

Using (2.4), and expanding the term $\left(1 - \exp\left(-\lambda/x\right)\right)^{\alpha(b+k)}$ we obtain

$$\begin{aligned} \mathrm{E}(|X-\mu|) &= \frac{2}{B(a,b)} \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \frac{(-1)^{(k+1+j)}}{a(b+k)B(a-k,k+1)\left[\alpha(b+k)+1\right]} \\ &\quad \times \frac{1}{B(j+1, \alpha(b+k)-j+1)} \int_0^{\mu} \exp\left(-j\lambda/x\right)\ dx\ , \end{aligned}$$

where

$$\int_0^c \exp(-j\lambda/x)\,dx = c\exp(-j\lambda/c) - j\lambda\Gamma(0, j\lambda/c)\,. \tag{3.11}$$

Hence, the theorem is proved. □

3.4.2. The mean deviation from the median can be found from the following theorem:

Theorem 2. *The mean deviation from the median of the BGIED is in the form*

$$\begin{aligned} E(|X-m|) &= \mu + \frac{2}{B(a,b)} \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} \frac{(-1)^{(k+j)} j\lambda}{a(b+k)B(a-k,k+1)\left[\alpha(b+k)+1\right]} \\ &\quad \times \frac{\Gamma(0, j\lambda/m)}{B(j+1, \alpha(b+k)-j+1)}\,, \qquad j\lambda > 0, \quad m > 0\,. \end{aligned}$$

Proof: The mean deviation from the median can be defined as

$$\begin{aligned} \mathrm{E}(|X-m|) &= \int_0^\infty |x-m| f(x)dx \\ &= 2\int_0^m (m-x)\,f(x)dx - \int_0^m (m-x)\,f(x)dx + \int_m^\infty (x-m)\,f(x)dx \\ &= 2\int_0^m (m-x)\,f(x)dx + \int_0^\infty (x-m)\,f(x)dx \qquad (3.12)\\ &= \mu - 2\left[mF(m) - \int_0^m F(x)dx\right] \\ &= \mu - m + 2\int_0^m F(x)dx\,. \end{aligned}$$

Substituting (2.4) into (3.12) and using (3.11), we obtain

$$\begin{aligned} \mathrm{E}(|X-m|) &= \mu - m \\ &\quad + \frac{2}{B(a,b)} \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} \frac{(-1)^{k+j+1}\left[m\exp(-j\lambda/m) - j\lambda\Gamma(0, j\lambda/m)\right]}{a(b+k)B(a-k,k+1)\left[\alpha(b+k)+1\right]} \\ &\quad \times \frac{1}{B\left[j+1, \alpha(b+k)-j+1\right]} \\ &= \mu - m + 2mF(m) \\ &\quad + \frac{2}{B(a,b)} \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} \frac{(-1)^{k+j} j\lambda\Gamma(0, j\lambda/m)}{a(b+k)B(a-k,k+1)\left[\alpha(b+k)+1\right]} \\ &\quad \times \frac{1}{B\left[j+1, \alpha(b+k)-j+1\right]}\,. \end{aligned}$$

Hence, the theorem is proved. □

3.5. The mode

The mode for the BGIED can be found by differentiating $f(x)$ with respect to x; thus, (2.2) gives

$$
\begin{aligned}
f'(x) \;&=\; f(x)\left\{\frac{-2}{x}+\frac{\lambda}{x^2}-(\alpha b-1)\left[1-\exp\left(-\lambda/x\right)\right]^{-1}\frac{\lambda}{x^2}\exp\left(-\lambda/x\right)\right.\\
&\quad+\;(a-1)\left[1-\left(1-\exp\left(-\lambda/x\right)\right)^{\alpha}\right]^{-1}\\
&\quad\times\;\left.\frac{\alpha\lambda}{x^2}\exp\left(-\lambda/x\right)\left(1-\exp\left(-\lambda/x\right)\right)^{\alpha-1}\right\}.
\end{aligned}
\tag{3.13}
$$

By equating (3.13) with zero, we get

$$
\begin{aligned}
&1-\frac{2x}{\lambda}+\left(\exp\left(-\lambda/x\right)-1\right)^{-1}\times\\
&\quad\times\left\{\alpha\left(a-1\right)\left[\left(1-\exp\left(-\lambda/x\right)\right)^{-\alpha}-1\right]^{-1}-(\alpha b-1)\right\} \;=\; 0\,.
\end{aligned}
\tag{3.14}
$$

Then, the mode of the BGIED can be found numerically by solving (3.14).

In Table 1, we present the values of the mean, standard deviation (SD), mode, median, skewness and kurtosis for different values of a, b, α and λ.

Table 1: The mean, SD, mode, median, skewness and kurtosis for different values of the parameters.

a	b	α	λ	mean	SD	mode	median	skewness	kurtosis
1	1	4	2	1.35919	1.04298	0.76393	1.08802	0.23016	0.66022
1	1	4	4	2.71838	2.08595	1.52787	2.17604	0.23016	0.66022
1	2	4	4	1.79735	0.88156	1.30871	1.60709	0.16158	0.44815
1	3	4	4	1.50143	0.61555	1.19585	1.38881	0.13109	0.35990
1	2	4	2	0.89867	0.44078	0.65435	0.80354	0.16158	0.44815
2	1	4	2	1.81971	1.24786	1.12748	1.50317	0.22047	0.63372
2	2	2	2	1.98654	1.39808	1.19879	1.62873	0.22323	0.63985
2	2	2	4	3.97308	2.79615	2.39758	3.25747	0.22323	0.63985
3.5	2	1	0.5	1.97910	302.436	0.65618	1.17708	0.32893	1.01650

The results of studying the behaviour of the BGIED are shown in Table 1 and Figure 1. We note that the distribution is unimodal and positively skewed. For fixed values of a, b and α, the kurtosis values remain constant; therefore, the mode, median and mean increase with the increase of λ. As we increase the value of $\alpha \geq 1$ and fix the other parameters, the kurtosis value increases and the mean decreases. It is noted that the distribution has a long right tail for fixed values of b, α and λ. Moreover, for fixed values of a, α and λ the kurtosis and the mean values decrease as we increase the value of b. Additionally, for different values of α and λ and fixed values of a and b, the skewness and the kurtosis values remain stable. Alternately, for fixed values of α and λ, the skewness and the kurtosis values decrease as we increase a and b. Furthermore, we found that our results for $a = b = 1$ are exactly the same as the results in Abouammoh and Alshingiti (2009).

4. MAXIMUM LIKELIHOOD ESTIMATORS

In this section, we examine estimation by maximum likelihood and inference for the BGIED. Let X_1, X_2, ..., X_n be a random sample from the BGIED with pdf and cdf given, respectively, by (2.2) and (2.4). The likelihood function in this case can be written as (Lawless (2003)):

$$L(\underline{\theta}|\underline{x}) = \prod_{i=1}^{n} f(x_i) \;, \tag{4.1}$$

where $f(\cdot)$ is given by (2.2) and $\underline{\theta} = (a, b, \alpha, \lambda)$.

The natural logarithm of the likelihood function (4.1) is given by

$$\ell = \log L(\underline{\theta}|\underline{x}) = \sum_{i=1}^{n} \log f(x_i) \;. \tag{4.2}$$

For the BGEID, we have have

$$\begin{aligned} \log L = \; & n \log(\alpha\lambda) - n \log B(a,b) + (\alpha b - 1) \sum_{i=1}^{n} \log\left(1 - \exp\left(\frac{-\lambda}{x_i}\right)\right) \\ & -\lambda \sum_{i=1}^{n} x_i^{-1} - 2 \sum_{i=1}^{n} \log(x_i) + (a-1) \sum_{i=1}^{n} \log\left[1 - \left(1 - \exp\left(\frac{-\lambda}{x_i}\right)\right)^{\alpha}\right] . \end{aligned} \tag{4.3}$$

Assuming that the parameters $\underline{\theta} = (a, b, \alpha, \lambda)$, are unknown, the likelihood equations are given for $\underline{\theta}$

$$l_j = \frac{\partial \log L}{\partial \theta_j} = \frac{1}{f(x_i)} \frac{\partial f(x_i)}{\partial \theta_j} = 0 \;, \qquad j = 1, 2, 3, 4 \;.$$

From (2.2), we have

$$\frac{\partial \log L}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n} \log\left(1 - \exp\left(\frac{-\lambda}{x_i}\right)\right) \left\{ b - (a-1) \left[\left(1 - \exp\left(\frac{-\lambda}{x_i}\right)\right)^{-\alpha} - 1\right]^{-1} \right\} , \tag{4.4}$$

$$\begin{aligned}\frac{\partial \log L}{\partial \lambda} &= \frac{n}{\lambda} + (\alpha b - 1)\sum_{i=1}^{n} x_i^{-1}\left(\exp\left(\frac{\lambda}{x_i}\right) - 1\right)^{-1} - \sum_{i=1}^{n} x_i^{-1} \\ &\quad -\alpha(a-1)\sum_{i=1}^{n} x_i^{-1}\left[\left(1-\exp\left(\frac{-\lambda}{x_i}\right)\right)^{-\alpha} - 1\right]^{-1}\left(\exp\left(\frac{\lambda}{x_i}\right) - 1\right)^{-1},\end{aligned} \tag{4.5}$$

$$\frac{\partial \log L}{\partial a} = \frac{-n}{B(a,b)}\phi_1 + \sum_{i=1}^{n}\log\left[1-\left(1-\exp\left(\frac{-\lambda}{x_i}\right)\right)^{\alpha}\right],$$

$$\begin{aligned}\phi_1 &= \frac{\partial B(a,b)}{\partial a} = \frac{\Gamma(b)\left[\Gamma(a+b)\Gamma'(a) - \Gamma(a)\partial\Gamma(a+b)/\partial a\right]}{\left[\Gamma(a+b)\right]^2} \\ &= B(a,b)\left[\psi(a) - \psi(a+b)\right],\end{aligned}$$

where $\psi(z) = \frac{1}{\Gamma(z)}\frac{\partial\Gamma(z)}{\partial z} = \frac{\Gamma'(z)}{\Gamma(z)}$ is called the Psi function (Abramowitz and Stegun (1972)). Then,

$$\frac{\partial \log L}{\partial a} = -n\left[\psi(a) - \psi(a+b)\right] + \sum_{i=1}^{n}\log\left[1-\left(1-\exp\left(\frac{-\lambda}{x_i}\right)\right)^{\alpha}\right], \tag{4.6}$$

$$\frac{\partial \log L}{\partial b} = \frac{-n}{B(a,b)}\phi_2 + \alpha\sum_{i=1}^{n}\log\left(1-\exp\left(\frac{-\lambda}{x_i}\right)\right),$$

$$\begin{aligned}\phi_2 &= \frac{\partial B(a,b)}{\partial b} = \frac{\Gamma(a)\left[\Gamma(a+b)\Gamma'(b) - \Gamma(b)\partial\Gamma(a+b)/\partial b\right]}{\left[\Gamma(a+b)\right]^2} \\ &= B(a,b)\left[\psi(b) - \psi(a+b)\right],\end{aligned}$$

$$\frac{\partial \log L}{\partial b} = -n\left[\psi(b) - \psi(a+b)\right] + \alpha\sum_{i=1}^{n}\log\left(1-\exp\left(\frac{-\lambda}{x_i}\right)\right). \tag{4.7}$$

The solution of the four nonlinear likelihood equations via (4.4), (4.5), (4.6) and (4.7) yields the maximum likelihood estimates (MLEs) $\underline{\hat{\theta}} = (\hat{a}, \hat{b}, \hat{\alpha}, \hat{\lambda})$ of $\underline{\theta} = (a, b, \alpha, \lambda)$. These equations are in implicit form, so they may be solved using numerical iteration, such as the Newton–Raphson method via Mathematica 9.0.

5. ASYMPTOTIC VARIANCES AND COVARIANCES OF ESTIMATES

The asymptotic variances of maximum likelihood estimates are given by the elements of the inverse of the Fisher information matrix $I_{ij}\left(\underline{\theta}\right) = E\left(-\frac{\partial^2 \ln L}{\partial\theta_i\partial\theta_j}\right)$. Unfortunately, the exact mathematical expressions for the above expectation are very difficult to obtain. Therefore, the observed Fisher information matrix is given by $I_{ij} = -\frac{\partial^2 \ln L}{\partial\theta_i \partial\theta_j}$ which is obtained by dropping the expectation on operation E (Cohen (1965)). The approximate (observed) asymptotic variance-covariance matrix F for the maximum likelihood estimates of the BGIED can be written as follows

$$F = [I_{ij}(\underline{\theta})] , \qquad i,j = 1,2,3,4 \quad \text{and} \quad \underline{\theta} = (a,b,\alpha,\lambda) .$$

The second partial derivatives of the maximum likelihood function for the BGIED are given as the following

$$\begin{aligned}
\frac{\partial^2 \log L}{\partial \lambda^2} &= \frac{-n}{\lambda^2} - (\alpha b - 1)\sum_{i=1}^{n} x_i^{-2}\left(\exp\left(\frac{\lambda}{x_i}\right)-1\right)^{-1}\left[1+\left(\exp\left(\frac{\lambda}{x_i}\right)-1\right)^{-1}\right] \\
&\quad - \alpha(a-1)\sum_{i=1}^{n} x_i^{-2}\left[\left(1-\exp\left(\frac{-\lambda}{x_i}\right)\right)^{-\alpha}-1\right]^{-1}\left(\exp\left(\frac{\lambda}{x_i}\right)-1\right)^{-1} \\
&\quad \times\left\{-1+\left(\exp\left(\frac{\lambda}{x_i}\right)-1\right)^{-1}\left((\alpha-1)+\alpha\left[\left(1-\exp\left(-\frac{\lambda}{x_i}\right)\right)^{-\alpha}-1\right]^{-1}\right)\right\} ,
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \log L}{\partial \alpha^2} &= \frac{-n}{\alpha^2} - (a-1)\sum_{i=1}^{n}\left[\log\left(1-\exp\left(\frac{-\lambda}{x_i}\right)\right)\right]^2\left[\left(1-\exp\left(\frac{-\lambda}{x_i}\right)\right)^{-\alpha}-1\right]^{-1} \\
&\quad \times\left\{1+\left[\left(1-\exp\left(\frac{-\lambda}{x_i}\right)\right)^{-\alpha}-1\right]^{-1}\right\} ,
\end{aligned}$$

$$\frac{\partial^2 \log L}{\partial a^2} = -n\left[\psi'(a)-\psi'(a+b)\right] ,$$

where

$$\begin{aligned}
\psi'(z) &= \frac{\partial \psi(z)}{\partial z} , \\
\frac{\partial^2 \log L}{\partial b^2} &= -n\left[\psi'(b)-\psi'(a+b)\right] , \\
\frac{\partial^2 \log L}{\partial a\, \partial b} &= n\,\psi'(a+b) ,
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \log L}{\partial \lambda\, \partial \alpha} &= \sum_{i=1}^{n} x_i^{-1}\left(\exp\left(\frac{\lambda}{x_i}\right)-1\right)^{-1}\left\{b-(a-1)\left[\left(1-\exp\left(\frac{-\lambda}{x_i}\right)\right)^{-\alpha}-1\right]^{-1}\right. \\
&\quad \left.\times\left[1+\frac{\alpha\log\left(1-\exp\left(\frac{-\lambda}{x_i}\right)\right)}{\left[1-\left(1-\exp\left(\frac{-\lambda}{x_i}\right)\right)^{\alpha}\right]}\right]\right\} ,
\end{aligned}$$

$$\frac{\partial^2 \log L}{\partial \lambda\, \partial a} = -\alpha\sum_{i=1}^{n} x_i^{-1}\left[\left(1-\exp\left(\frac{-\lambda}{x_i}\right)\right)^{-\alpha}-1\right]^{-1}\left(\exp\left(\frac{\lambda}{x_i}\right)-1\right)^{-1} ,$$

$$\frac{\partial^2 \log L}{\partial\lambda\,\partial b} = \alpha \sum_{i=1}^{n} x_i^{-1} \left(\exp\left(\frac{\lambda}{x_i}\right) - 1\right)^{-1},$$

$$\frac{\partial^2 \log L}{\partial\alpha\,\partial a} = -\sum_{i=1}^{n} \log\left(1 - \exp\left(\frac{-\lambda}{x_i}\right)\right)\left[\left(1 - \exp\left(\frac{-\lambda}{x_i}\right)\right)^{-\alpha} - 1\right]^{-1},$$

$$\frac{\partial^2 \log L}{\partial\alpha\,\partial b} = \sum_{i=1}^{n} \log\left(1 - \exp\left(\frac{-\lambda}{x_i}\right)\right).$$

Consequently, the maximum likelihood estimators of a, b, α and λ and have an asymptotic variance-covariance matrix defined by inverting the Fisher information matrix F and by substituting $\hat{a}$ for a, $\hat{b}$ for b, $\hat{\alpha}$ for α and $\hat{\lambda}$ for λ.

6. INTERVAL ESTIMATES

If $L_\theta = L_\theta\,(y_1, ..., y_n)$ and $U_\theta = U_\theta\,(y_1, ..., y_n)$ are functions of the sample data $y_1, ..., y_n$ then a confidence interval for a population parameter θ is given by

$$P\Big(L_\theta \le \theta \le U_\theta\Big) = \gamma\,,$$

where L_θ and U_θ are the lower and upper confidence limits that enclose θ with probability γ. The interval $[L_\theta, U_\theta]$ is called a $100\gamma\%$ confidence interval for θ.

For large sample sizes (Bain and Engelhardt (1992)), the maximum likelihood estimates, under appropriate regularity conditions, are consistent and asymptotically normally distributed. Therefore, the approximate $100\gamma\%$ confidence limits for the maximum likelihood estimate $\hat{\theta}$ of a population parameter θ can be constructed, such that

$$P\left(-z \le \frac{\hat{\theta} - \theta}{\sigma(\hat{\theta})} \le z\right) = \gamma\,, \tag{6.1}$$

where z is the $\frac{100(1+\gamma)}{2}$ standard normal percentile. Therefore, the approximate $100\gamma\%$ confidence limits for a population parameter θ can be obtained such that

$$P\left(\hat{\theta} - z\sigma(\hat{\theta}) \le \theta \le \hat{\theta} + z\sigma(\hat{\theta})\right) = \gamma\,. \tag{6.2}$$

Then, the approximate confidence limits for a, b, α and λ will be constructed using (6.2) with a confidence level of 90%.

7. SIMULATION STUDIES

Simulation studies have been performed using Mathematica 9.0 to illustrate the theoretical results of the estimation problem. The performance of the resulting estimators of the parameters has been considered in terms of their absolute relative bias (ARBias) and mean square error (MSE), where

$$\text{ARBias}\left(\hat{\theta}\right) = \left|\frac{\hat{\theta}-\theta}{\theta}\right| \qquad \text{and} \qquad \text{MSE}\left(\hat{\theta}\right) = \text{E}\left(\hat{\theta}-\theta\right)^2.$$

Furthermore, the asymptotic variance, covariance matrix and confidence intervals of the parameters are obtained. The algorithm for the simulation procedure is described below:

Step 1. 1000 random samples of sizes $n = 10(10)50, 100, 200$ and 300 were generated from the BGIED. The true parameter values are selected as $(a = 1,\ b = 2,\ \alpha = 4,\ \lambda = 2)$.

Step 2. For each sample, the parameters of the distribution are estimated under the complete sample.

Step 3. The Newton–Raphson method is used for solving the four nonlinear likelihoods for α, λ, a and b given in (4.4), (4.5), (4.6) and (4.7), respectively.

Step 4. The ARBiase and MSE of the estimators for the four parameters for all sample sizes are tabulated.

Step 5. For large sample sizes $n = 100, 200$ and 300, the Fisher information matrix of the estimators are computed using the equations presented in Section 5.

Step 6. By inverting the Fisher information matrix that was computed in Step 5, the asymptotic variances and covariances of the estimators are found.

Step 7. Based on the values of the asymptotic variances and covariances matrix that were found in Step 6 and on Eq. (6.2), the approximate confidence limits at 90% for the parameters are computed.

Simulation results are summarized in Tables 2, 3 and 4. Table 2 gives the ARBias and MSE of the estimators. The asymptotic variances and covariances matrix of the estimators for complete samples of size $n = 100, 200$ and 300 and true parameter values $(a = 1,\ b = 2,\ \alpha = 4,\ \lambda = 2)$ are displayed in Table 3. The approximate confidence limits at 90% for the parameters are presented in Table 4.

Table 2: The ARBias and MSE of the parameters $\underline{\theta} = (a, b, \alpha, \lambda)$.

n	$\hat{a}$		$\hat{b}$		$\hat{\alpha}$		$\hat{\lambda}$	
	ARBias	MSE	ARBias	MSE	ARBias	MSE	ARBias	MSE
10	2.77997	57.22180	1.23642	39.93170	0.34929	57.85210	0.04445	4.49668
20	1.57487	18.92080	0.75400	18.13730	0.01542	23.75870	0.02126	3.08791
30	1.17703	8.04064	0.59348	10.72860	0.13843	12.21520	0.03641	2.48411
40	1.01435	7.53433	0.32958	4.15377	0.25344	7.47854	0.12458	1.52254
50	0.81716	5.63524	0.35502	5.58289	0.23793	8.44674	0.08708	1.65372
100	0.34507	2.01241	0.13037	1.57136	0.37237	5.26341	0.15154	0.73170
200	0.09587	0.78450	0.04185	0.88988	0.45129	4.69708	0.18296	0.46773
300	0.05260	0.56916	0.04620	0.88359	0.46429	4.67339	0.18716	0.31603

Table 3: Asymptotic variances and covariances of estimates for complete samples.

n	Parameters	$\hat{a}$	$\hat{b}$	$\hat{\alpha}$	$\hat{\lambda}$
100	a	0.00860	0.00389	−0.00021	−0.00359
	b	0.00389	0.05006	−0.00513	0.00867
	α	−0.00021	−0.00513	0.04464	0.01414
	λ	−0.00359	0.00867	0.01414	0.01308
200	a	0.00673	0.00327	−0.00018	−0.00411
	b	0.00327	0.03238	−0.00883	0.00346
	α	−0.00018	−0.00883	0.02928	0.00772
	λ	−0.00411	0.00346	0.00772	0.00951
300	a	0.00630	0.00320	−0.00075	−0.00445
	b	0.00320	0.02563	−0.00722	0.00197
	α	−0.00075	−0.00722	0.02245	0.00664
	λ	−0.00445	0.00197	0.00664	0.00888

Table 4: Confidence bounds of the estimates at a confidence level of 0.90.

n	Parameters	Estimated mean	Lower bound	Upper bound	Width
100	a	1.34507	1.19294	1.49719	0.30424
	b	2.26075	1.89380	2.62770	0.73390
	α	2.51052	2.16399	2.85704	0.69304
	λ	1.69693	1.50931	1.88455	0.375241
200	a	1.09587	0.96132	1.23041	0.26909
	b	2.08370	1.78856	2.37883	0.59027
	α	2.19482	1.91416	2.47549	0.56132
	λ	1.63408	1.47409	1.79408	0.31999
300	a	1.05260	0.92241	1.18279	0.26038
	b	2.09239	1.82981	2.35498	0.52517
	α	2.14283	1.89709	2.38857	0.49147
	λ	1.62569	1.47108	1.78030	0.30922

From these tables, the following observations can be made on the performance of the parameter estimation of the BGIED:

1. As the sample size increases, the MSEs of the estimated parameters decrease. This indicates that the maximum likelihood estimates provide asymptotically normally distributed and consistent estimators for the parameters (see Table 2).

2. Although the estimators of a and b are consistent according to the ARBias, it is noted that the estimators of α and λ are not consistent. Table 2 shows that, for large sample sizes ($n = 100$, 200 and 300), the ARBiases are increased, which indicates that the estimates of α and λ are not consistent.

3. The asymptotic variances of the estimators decrease when the sample size is increasing (see Table 3).

4. The interval of the estimators decreases when the sample size is increasing (see Table 4).

5. The interval estimations of all parameters were reasonable except the interval estimate of α. The estimated intervals at a confidence level of 0.90 for $n = 100, 200$ and 300 did not cover the real value of α.

We conclude from the previous points that the MLE of the parameters is a good estimator for a, b and λ.

8. APPLICATIONS

In this section, two sets of data are presented to demonstrate the utility of using the BGIED. These two sets were investigated by Abouammoh and Alshingiti (2009). The GIED was fitted to both of these sets.

8.1. The first data-set

The following data-set is presented in Lawless (2003). The data resulted from a test on the endurance of deep groove ball bearings. The data are as follows:

17.88, 28.92, 33.0, 41.52, 42.12, 45.60, 48.40, 51.84, 51.96, 54.12, 55.56, 67.80, 68.64, 68.64, 68.88, 84.12, 93.12, 98.64, 105.12, 105.84, 127.92, 128.04, 173.4 .

Descriptive statistics of these data are tabulated in Table 5.

Table 5: Descriptive statistics for the ball bearing data.

Measure	Value	Measure	Value
n	23	Minimum	17.880
Maximum	173.400	Mean	72.22
$Q1$	45.600	$Q3$	98.640
Median	67.800	Mean deviation	29.429
Variance	1405.580	SD	37.491
Skewness	0.941	Kurtosis	3.486

We apply the Kolmogorov–Smirnov (K-S) statistic to verify which distribution better fits these data. The K-S test statistic is described in detail in D'Agostino and Stephens (1986). In general, the smaller the value of K-S is, the better the fit to the data is. All graphs and computations presented to analyse the data were carried out by Mathematica 9.0. The model selection was carried out using the AIC (Akaike information criterion), the BIC (Bayesian information criterion) and the CAIC (consistent Akaike information criterion):

$$AIC = -2\,l(\underline{\hat{\theta}}) + 2p\,,$$

$$BIC = -2\,l(\underline{\hat{\theta}}) + p\log n$$

and

$$CAIC = -2\,l(\underline{\hat{\theta}}) + \frac{2pn}{n-p-1}\,,$$

where $l(\underline{\hat{\theta}})$ denotes the log likelihood function evaluated at the maximum likelihood estimates, p is the number of parameters and n is the sample size. Table 6 lists the values of the K-S statistic and of $-2l(\underline{\hat{\theta}})$. The K-S goodness-of-fit test of the BGIED as well as the GIED are the best among all models; accordingly, the BGIED model can be used to analyse the ball bearing data. Table 7 provides the MLEs with corresponding standard errors (SEs) of the model parameters.

Table 6: Goodness-of-fit measures and K-S statistics for the ball bearing data.

Model	BGIED	GIED	IED	BIED
K-S statistics	0.097	0.091	0.306	0.104
$-2l(\underline{\hat{\theta}})$	227.723	227.098	243.452	228.288

Table 7: MLEs of the model parameters, the corresponding SEs and the statistics of the AIC, BIC and CAIC for the ball bearing data.

Model	Method	Estimates				Statistics		
		$\hat{a}$	$\hat{b}$	$\hat{\alpha}$	$\hat{\lambda}$	AIC	BIC	CAIC
BGIED	MLE	20.615	9.427	0.532	7.161	235.723	240.265	237.946
	SE	0.125	0.279	8.052	0.262			
GIED	MLE			5.307	129.996	231.098	233.369	231.698
	SE			0.188	0.015			
IED	MLE				55.055	245.452	246.587	245.642
	SE				0.018			
BIED	MLE	15.858	3.692		11.858	234.288	237.695	235.552
	SE	0.112	0.508		0.161			

8.2. The second data-set

The data that is studied in this section was provided by Ed Fuller of the NICT Ceramics Division in December 1993. It contains polished window strength data. Fuller *et al.* (1994) described the use of this set to predict the lifetime for a glass airplane window. The data are as follows:

18.83, 20.8, 21.657, 23.03, 23.23, 24.05, 24.321, 25.5, 25.52, 25.8, 26.96, 26.77, 26.78, 27.05, 27.67, 29.9, 31.11, 33.2, 33.73, 33.76, 33.89, 34.76, 35.75, 35.91, 36.98, 37.08, 37.09, 39.58, 44.045, 45.29, 45.381 .

Descriptive statistics of the window strength data are tabulated in Table 8.

Table 8: Descriptive statistics for the window strength data.

Measure	Value	Measure	Value
n	31	Minimum	18.830
Maximum	45.381	Mean	30.820
$Q1$	25.500	Q3	35.910
Median	29.900	Mean deviation	6.145
Variance	52.539	SD	7.248
Skewness	0.403	Kurtosis	2.290

The K-S goodness-of-fit test of the BGIED as well as the BIED are the best among all models; therefore, the BGIED model can be used to study the window strength data. Table 10 presents the MLEs with corresponding SEs of the model parameters.

Table 9: Goodness-of-fit measures and K-S statistics for the window strength data.

Model	BGIED	GIED	IED	BIED
K-S statistics	0.133	0.137	0.474	0.130
$-2l(\hat{\underline{\theta}})$	208.207	208.454	274.523	208.105

Table 10: MLEs of the model parameters, the corresponding SEs and the statistics of the AIC, BIC and CAIC for the window strength data.

Model	Method	Estimates				Statistics		
		$\hat{a}$	$\hat{b}$	$\hat{\alpha}$	$\hat{\lambda}$	AIC	BIC	CAIC
BGIED	MLE	27.850	7.354	1.978	17.601	216.207	221.943	217.745
	SE	0.088	0.341	2.401	0.247			
GIED	MLE			90.855	148.412	212.454	215.322	212.883
	SE			0.011	0.029			
IED	MLE				29.215	276.523	277.957	276.661
	SE				0.034			
BIED	MLE	14.506	20.169		26.053	214.105	218.407	214.994
	SE	0.205	0.146		0.166			

Finally, we conclude the following from studying the AIC, BIC and CAIC statistics of the two previous data-sets.

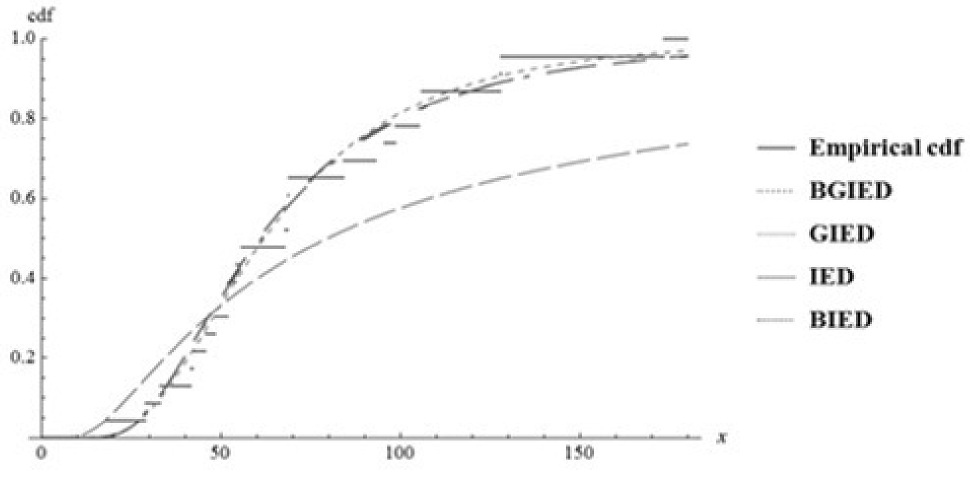


Figure 4: The empirical distribution and estimated cdf of the models for the ball bearing data.

It is noted that the GIED has a smaller value compared with the values of other models for the two data-sets. The BIED and BGIED follow next. That indicates that the GIED seems to be a very competitive model for these data. Because the values of the AIC, BIC and CAIC are approximately equivalent for the GIED, BIED and BGIED, the BGIED can thus be a good alternative model for these data, as can the GIED. Alternately, the IED presents the worst fit for the second dataset. Figure 4 shows the empirical distribution and estimated cdf of the models for the ball bearing data. Figure 5 shows the empirical distribution and estimated cdf of the models for the window strength data.

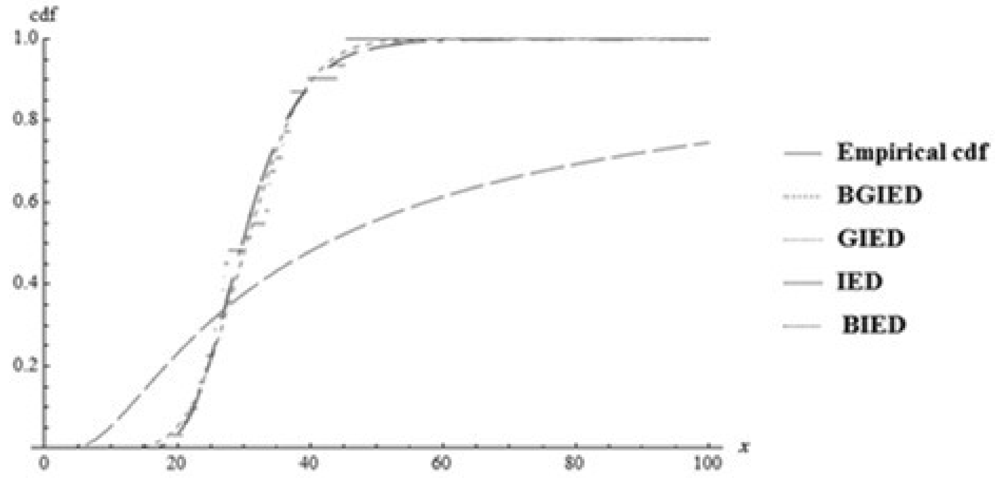


Figure 5: The empirical distribution and estimated cdf of the models for the window strength data.

9. CONCLUDING REMARKS

In this study, the four-parameter beta generalized inverted exponential distribution (BGIED) is proposed. BGIED generalizes the generalized inverted exponential distribution discussed by Abouammoh and Alshingiti (2009). Additionally, the BGIED represents a generalization of the inverted exponential distribution (IED). IED has been considered by Keller and Kamath (1982) and Duran and Lewis (1989). Statistical properties of the BGIED are studied. Maximum likelihood estimators of the BGIED parameters are obtained. The information matrix and the asymptotic confidence bounds of the parameters are derived. Monte Carlo simulation studies are conducted under different sample sizes to study the theoretical performance of the MLE of the parameters. Two real data-sets are analysed, and the BGIED has provided a good fit for the data-sets.

ACKNOWLEDGMENTS

The authors would like to thank the referees for their useful comments and suggestions. Additionally, the authors acknowledge with thanks the Deanship of Scientific Research, King Abdulaziz University, Jeddah, for the technical and financial support.

REFERENCES

[1] Abouammoh, A. M. and Alshingiti, A. M. (2009). Reliability estimation of generalized inverted exponential distribution, *Journal of Statistical Computation and Simulation*, **79**(11), 1301–1315.

[2] Abramowitz, M. and Stegun, I. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, 2nd Ed., Dover Publications, New York.

[3] Baharith, L. A.; Mousa, S. A.; Atallah, M. A. and Elgayar, S. H. (2014). The beta generalized inverse Weibull distribution, *British Journal of Mathematics and Computer Science*, **4**(2), 252–270.

[4] Bain, L. J. and Engelhardt, M. (1992). *Introduction to Probability and Mathematical Statistics*, 2nd Ed., Duxbury Press, Boston.

[5] Barreto-Souza, W.; Santos, A. H. S. and Cordeiro, G. M. (2010). The beta generalized exponential distribution, *Journal of Statistical Computation and Simulation*, **80**, 159–172.

[6] Cohen, A. C. (1965). Maximum likelihood estimation in the Weibull distribution based on complete and censored samples, *Technometrics*, **5**, 579–588.

[7] Cordeiro, G. M. and Lemonte, A. J. (2011). The beta Laplace distribution, *Statistics and Probability Letters*, **81**, 973–982.

[8] D'Agostino, R. B. and Stephens, M. A. (1986). *Goodness-of-Fit Techniques*, Marcel Dekker, New York.

[9] Duran, B. S. and Lewis, T. O. (1989). Inverted gamma as life distribution, *Microelectron. Reliab.*, **29**, 619–626.

[10] Eugene, N.; Lee, C. and Famoye, F. (2002). Beta-normal distribution and its applications, *Commun. Statist. – Theory and Methods*, **31**, 497–512.

[11] Fuller Jr, E. R.; Frieman, S. W.; Quinn, J. B.; Quinn, G. D. and Carter, W. C. (1994). Fracture mechanics approach to the design of glass aircraft windows: A case study, *SPIE Proc.*, **2286**, 419–430.

[12] Gupta, A. K. and Nadarajah, S. (2004). On the moments of the beta normal distribution, *Commun. Statist. – Theory and Methods*, **33**, 1–13.

[13] Keller, A. Z. and Kamath, A. R. (1982). Reliability analysis of CNC Machine Tools, *Reliab. Eng.*, **3**, 449–473.

[14] Kenney, J. F. and Keeping, E. S. (1962). *Mathematics of Statistics*, 3rd Ed., Part 1, Princeton, New Jersey.

[15] Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*, 2nd Ed., Wiley, New York.

[16] Leão, J.; Saulo, H.; Bourguignon, M.; Cintra, R.; Rego, L. and Cordeiro, G. (2013). On some properties of the beta inverse Rayleigh distribution, *Chilean Journal of Statistics*, **4**(2), 111–131.

[17] Lee, C.; Famoye, F. and Olumolade, O. (2007). The beta-Weibull distribution: Some properties and applications censored data, *J. Modern Applied Statistical Method*, **6**(1), 173–186.

[18] Mahmoudi, E. (2011). The beta generalized Pareto distribution with application to lifetime data, *Mathematics and Computers in Simulation*, **81**, 2414–2430.

[19] Moors, J. J. (1988). A quantile alternative for kurtosis, *Journal of the Royal Statistical Society D*, **37**, 25–32.

[20] Nadarajah, S. and Kotz, S. (2005). The beta exponential distribution, *Reliability Engineering and System Safety*, **91**, 689–697.

[21] Nadarajah, S. and Kotz, S. (2004). The beta Gumbel distribution, *Math. Probab. Eng.*, **10**, 323–332.

[22] Nassar, M. M. and Nada N. K. (2011). The beta generalized Pareto distribution, *Journal of Statistics: Advances in Theory and Applications*, **6**(1/2), 1–17.

[23] Paranaiba, P. F.; Ortega, E. M. M.; Cordeiro, G. M. and Pescim, R. P. (2011). The beta Burr XII distribution with application to lifetime data, *Computational Statistics and Data Analysis*, **55**(2), 1118–1136.

[24] Zea, L. M.; Silva, R. B.; Bourguignon, M.; Santos, A. M. and Cordeiro, G.M. (2012). The Beta exponentiated Pareto distribution with application to bladder cancer susceptibility, *International Journal of Statistics and Probability*, **1**(2), 8–19.

REVSTAT – Statistical Journal
Volume 15, Number 1, January 2017, 89–105

SOME CHARACTERIZATION RESULTS BASED ON DOUBLY TRUNCATED REVERSED RESIDUAL LIFETIME RANDOM VARIABLE

Authors: M. Khorashadizadeh
– Department of Statistics, University of Birjand, Iran
m.khorashadizadeh@birjand.ac.ir

A.H. Rezaei Roknabadi
– Department of Statistics, Ferdowsi University of Mashhad, Iran
rezaei@um.ac.ir

G.R. Mohtashami Borzadaran
– Department of Statistics, Ferdowsi University of Mashhad, Iran
grmohtashami@um.ac.ir

Received: December 2014 Revised: March 2015 Accepted: May 2015

Abstract:

- In view of the growing importance of the reversed residual lifetime in reliability analysis and stochastic modeling, in this paper, we try to study some of the reliability properties of reversed residual lifetime random variable based on doubly truncated data and complete the preceding results such as Ruiz and Navarro (1995, 1996), Navarro *et al.* (1998), Nanda *et al.* (2003), Nair and Sudheesh (2008) and Sudheesh and Nair (2010). The monotonicity properties of the doubly truncated reversed residual variance and its relations with doubly truncated reversed residual expected value and doubly truncated reversed residual coefficient of variation are discussed. Furthermore, an upper bound for it under some conditions is obtained. We also discuss and find the similar results and some characterizations for discrete random ageing, which are noticeable in comparing with continuous cases.

Key-Words:

- *doubly truncated reversed residual expected value; doubly truncated reversed residual coefficient of variation; doubly truncated reversed residual variance; generalized failure rate; reversed failure rate.*

AMS Subject Classification:

- 62N05, 90B25.

1. INTRODUCTION

In the reliability literature and analyzing survival data for the components of a system or a device, there have been defined several measures for various conditions and situations of the system. Sometimes, we only have information about between two lifetime points, so studying the reliability measures under the condition of doubly truncated random variables, is necessary. If the random variable X denotes the lifetime of a unit, then the random variable ${}_xX_y = (y - X \mid x \le X \le y)$ is called the doubly truncated reversed residual lifetime and it measures the time elapsed since the failure of X given that the system has been working since time x and has failed sometime before y. Note that the well-known random variable, $X_y = (y - X \mid X \le y)$, which some of researchers called it, "reversed residual lifetime (RRL)", "past time to failure", "inactivity time" or "idle time", is the special case of ${}_xX_y$ when $x = 0$.

The subject of doubly truncation of a lifetime random variable in reliability literature has been started by Navarro and Ruiz (1996) and Ruiz and Navarro (1995, 1996) that generalized the failure rate function for doubly truncated random variables. Later, Sankaran and Sunoj (2004) defined and obtained some properties of the expected value of the doubly truncated lifetime distributions. Recently, many authors such as Su and Huang (2000), Ahmad (2001), Betensky and Martin (2003), Navarro and Ruiz (2004), Bairamov and Gebizlioglu (2005), Poursaeed and Nematollahi (2008) and Sunoj *et al.* (2009), studied the properties of the conditional expectations of doubly truncated random variables in various areas like order statistics and k-out-of-n systems. Also, recently Khorashadizadeh *et al.* (2012) have studied the doubly truncated mean residual lifetime and the doubly truncated mean past to failure and also obtained some characterization results in both continuous and discrete cases.

In this paper, we study some reliability measures based on the doubly truncated reversed lifetime random variable, ${}_xX_y$, in both continuous and discrete lifetime distributions, which some of the results that achieved are not similar to each other. The relationship among doubly truncated reversed residual expected (mean) value ($dRRM$), doubly truncated reversed residual variance ($dRRV$) and doubly truncated reversed residual coefficient of variation ($dRRCV$) are obtained. Also, their monotonicity and the associated ageing classes of distributions are discussed. Some characterization results of the class of the increasing $dRRV$ are presented and an upper bound for $dRRV$ under some conditions is obtained. Furthermore, we characterize the discrete distribution based on doubly truncated covariance and obtained some results for binomial, Poisson and negative binomial distributions.

2. CONTINUOUS DOUBLY TRUNCATED REVERSED RESIDUAL LIFETIME

Let X be a non-negative continuous random variable with cumulative distribution function (cdf), $F(x)$ and probability density function (pdf), $f(x)$. Navarro and Ruiz (1996) defined and studied the generalized failure rate (GFR) to the doubly truncated continuous random variables by

$$h_1(x,y) = \lim_{h\to 0^+}\left[\frac{P\big(x \le X \le x+h \mid x \le X \le y\big)}{h}\right] = \frac{f(x)}{F(y)-F(x)}$$

and

$$h_2(x,y) = \lim_{h\to 0^-}\left[\frac{P\big(y+h \le X \le y \mid x \le X \le y\big)}{h}\right] = \frac{f(y)}{F(y)-F(x)}\ ,$$

for $(x,y) \in D = \{(x,y);\ F(x) < F(y)\}$. Note that the special cases, $h_1(x,\infty) = \frac{f(x)}{1-F(x)}$ is the failure rate and $h_2(0,y) = \frac{f(y)}{F(y)}$ is the reversed failure rate. Navarro and Ruiz (1996) have shown that GFR determines the distribution uniquely.

We denote the continuous doubly truncated reversed residual expected value by *dRRM* and the continuous doubly truncated reversed residual variance by *dRRV* and define them as

$$\tilde{\mu}(x,y) = E({}_xX_y) = E\big(y-X \,|\, x \le X \le y\big)$$

and

$$\tilde{\sigma}^2(x,y) = \mathrm{Var}({}_xX_y) = \mathrm{Var}\big(y-X \,|\, x \le X \le y\big) = \mathrm{Var}\big(X \,|\, x \le X \le y\big)\ ,$$

respectively, such that $E(X^2) < \infty$, $(x,y) \in D$ and $\tilde{\mu}(x,x) = \tilde{\sigma}^2(x,x) = 0$. Ruiz and Navarro (1995, 1996) and Navarro *et al.* (1998) have shown that $m(x,y) = E(X \,|\, x \le X \le y)$ determines $F(x)$ uniquely. So, this is also true for $\tilde{\mu}(x,y) = y - m(x,y)$.

The *dRRM* can be rewritten as

$$\begin{aligned} \tilde{\mu}(x,y) &= E\big(y-X \,|\, x \le X \le y\big) \\ &= \frac{(x-y)\,F(x) + \int_x^y F(t)\,dt}{F(y)-F(x)}\ . \end{aligned} \tag{2.1}$$

So, we have

$$\frac{\partial}{\partial y}\tilde{\mu}(x,y) = \frac{\big[F(y)-F(x)\big]^2 - f(y)\big[(x-y)\,F(x) + \int_x^y F(t)\,dt\big]}{\big[F(y)-F(x)\big]^2}\ .$$

By using the above equation, the $\tilde{\mu}(x,y)$ determine the general failure rate, $h_2(x,y)$, via relation

$$h_2(x,y) = \frac{1-\frac{\partial}{\partial y}\tilde{\mu}(x,y)}{\tilde{\mu}(x,y)}\ , \qquad (x,y) \in D\ .$$

Furthermore, using part by part integration method, we can see that $\tilde{\sigma}^2(x,y)$ and $\tilde{\mu}(x,y)$ are related via the following equation:

$$\tilde{\sigma}^2(x,y) \;=\; \frac{(y^2-x^2)\,F(x) - 2\int_x^y tF(t)\,dt}{F(y)-F(x)} + 2y\,\tilde{\mu}(x,y) - \tilde{\mu}^2(x,y)\;. \tag{2.2}$$

This equation will be useful in proving various other relationships. We have the following definitions of ageing classes related to the $\tilde{\mu}(x,y)$ and $\tilde{\sigma}^2(x,y)$.

Definition 2.1. A random variable X is said to be

(**i**) increasing in doubly truncated reversed residual expected value ($IdRRM$) if for any $(x,y)\in D$, $\tilde{\mu}(x,y)$ is increasing in y,

(**ii**) increasing in doubly truncated reversed residual variance ($IdRRV$) if for any $(x,y)\in D$, $\tilde{\sigma}^2(x,y)$ is increasing in y.

The dual classes are defined similarly. For the random variable X_y, Nanda *et al.* (2003) showed that the class of decreasing RRM is empty. Thus, in the next theorem, we answer the natural question that whether the classes of decreasing doubly truncated reversed residual expected value ($DdRRM$) and decreasing doubly truncated reversed residual variance ($DdRRV$) of life distributions are null or not.

Theorem 2.1.

I. *There exist no non-negative random variable that has DdRRM property.*

II. *There exist no non-negative random variable that has DdRRV property.*

Proof: The two part can be proved by assuming the opposite. Suppose that $\tilde{\mu}(x,y)$ is decreasing in y. From (2.1), we have

$$0 \le \tilde{\mu}(x,y) \le y-x\;, \qquad \forall (x,y)\in D\;,$$

and also

$$\lim_{y\to x} \tilde{\mu}(x,y) = 0\;.$$

Thus, if $\tilde{\mu}(x,y)$ is decreasing in y, then $\tilde{\mu}(x,y)\le\tilde{\mu}(x,x)=0$, for all $(x,y)\in D$, which is contradict the fact that $\tilde{\mu}(x,y)$ cannot be negative or identically zero. Similarly, for part II., on contrary, suppose that $\tilde{\sigma}^2(x,y)$ is decreasing in y. For all $(x,y)\in D$, we have

$$0 \le \tilde{\sigma}^2(x,y) \le (y^2-x^2)$$

and

$$\lim_{y\to x} \tilde{\sigma}^2(x,y) = 0\;.$$

Thus, if $\tilde{\sigma}^2(x,y)$ is decreasing in y, then $\tilde{\sigma}^2(x,y)\le\tilde{\sigma}^2(x,x)=0$, for all $(x,y)\in D$, which is contradict the fact that variance cannot be negative or identically zero. $\square$

In the next theorem, we obtain an upper bound for $\tilde{\sigma}^2(x,y)$, when X has the $IdRRM$ property.

Theorem 2.2. *If the non-negative continuous random variable X, has the $IdRRM$ property, then,*

$$\tilde{\sigma}^2(x,y) < \tilde{\mu}^2(x,y)\ , \qquad (x,y) \in D\ . \tag{2.3}$$

Proof: According to (2.1), we have

$$\begin{aligned}\int_x^y \big[F(t)-F(x)\big]\,\tilde{\mu}(x,t)\,dt \;&=\; \int_x^y\left[\int_x^t F(z)\,dz + (x-t)\,F(x)\right]dt\\ &=\; y\int_x^y F(z)\,dz - \int_x^y z\,F(z)\,dz + \int_x^y (x-t)\,F(x)\,dt\ ,\end{aligned}$$

using (2.2), it implies that

$$\begin{aligned}\frac{2}{F(y)-F(x)}\int_x^y \big[F(t)-F(x)\big]\,\tilde{\mu}(x,t)\,dt \;&=\\ &=\; \frac{(y^2-x^2)\,F(x) - 2\int_x^y zF(z)\,dz}{F(y)-F(x)} + 2\,y\,\tilde{\mu}(x,y)\\ &=\; \tilde{\mu}^2(x,y) + \tilde{\sigma}^2(x,y)\ .\end{aligned}$$

So, we have

$$\begin{aligned}\tilde{\sigma}^2(x,y) - \tilde{\mu}^2(x,y) \;&=\; \frac{2}{F(y)-F(x)}\int_x^y \big[F(t)-F(x)\big]\,\big[\tilde{\mu}(x,t)-\tilde{\mu}(x,y)\big]\,dt\\ &<\; 0\ ,\end{aligned}$$

since $\tilde{\mu}(x,y)$ is increasing in y. This completes the proof. □

Now, we investigate the connection between $IdRRV$ and other classes of life distributions.

Theorem 2.3. *If $\tilde{\mu}(x,y)$ is increasing in y, then $\tilde{\sigma}^2(x,y)$ is increasing in y, i.e., the $IdRRM$ property is stronger than the $IdRRV$ property.*

Proof: The proof is trivial by using the following relation:

$$\frac{\partial}{\partial y}\tilde{\sigma}^2(x,y) \;=\; h_2(x,y)\,\big[\tilde{\mu}^2(x,y) - \tilde{\sigma}^2(x,y)\big]\ . \tag{2.4}$$ □

In special case, the Example 2.1 in Nanda *et al.* (2003) shows that the converse of the above theorem is not true.

Another reliability measure that has been recently considered and is related to the reversed residual variance and the reversed residual expected value, is the reversed residual coefficient of variation. So, in doubly truncated random variables we consider the doubly truncated reversed residual coefficient of variation ($dRRCV$) as

$$\tilde{\gamma}(x,y) = \frac{\tilde{\sigma}(x,y)}{\tilde{\mu}(x,y)}\,, \qquad (x,y) \in D\,. \tag{2.5}$$

The Eq. (2.4) can be written as

$$\frac{\partial}{\partial y}\tilde{\sigma}^2(x,y) \;=\; h_2(x,y)\,\tilde{\mu}^2(x,y)\left[1-\tilde{\gamma}^2(x,y)\right]\,, \tag{2.6}$$

so, $\tilde{\sigma}^2(x,y)$ is increasing in y according as $\tilde{\gamma}^2(x,y) \le 1$.

The next theorem characterizes the monotonic behavior of the variance of the random variable ${}_xX_y$. A similar result for the variance of ${}_xX$ has been given by Nanda *et al.* (2003).

Theorem 2.4. *The following statements are equivalent:*

(i) $\tilde{\sigma}^2(x,y)$ *is increasing in* y *for any fixed* x *such that* $(x,y) \in D$.

(ii) $\tilde{\gamma}^2(x,y) \le 1$ *for all* $(x,y) \in D$.

(iii) $\Phi(x,y) = \frac{E\left[(y-X)^2 \,|\, x \le X \le y\right]}{E\left[y-X \,|\, x \le X \le y\right]}$ *is increasing in* y *for any fixed* x *such that* $(x,y) \in D$.

Proof: Using (2.6) and (2.2) the results will follow. □

3. DISCRETE DOUBLY TRUNCATED REVERSED RESIDUAL LIFETIME

In reliability analysis, interests in discrete failure data came relatively late in comparison to its continuous analogue.

Suppose T be a non-negative discrete random variable with support $\{0,1,2,...\}$ and cdf, $F(t)$ and probability mass function (pmf), $p(t)$. Navarro and Ruiz (1996) defined the generalized failure rate (GFR) to the doubly truncated discrete random variables for all $(t,k) \in D^* = \{(t,k);\, F(t^-) < F(k)\}$ by

$$h_1(t,k) \;=\; \frac{p(t)}{F(k)-F(t-1)} \tag{3.1}$$

and

$$h_2(t,k) \;=\; \frac{p(k)}{F(k)-F(t-1)}\,. \tag{3.2}$$

Let ${}_tT_k = (k-T \,|\, t \le T \le k)$ be the reversed doubly truncated random variable in discrete lifetime distributions. So, the doubly truncated reversed residual expected value and reversed residual variance based on ${}_tT_k$ are as follow,

$$\begin{aligned} \tilde{\mu}_d(t,k) &= E({}_tT_k) = E(k-T \,|\, t \le T \le k) , \\ \tilde{\sigma}_d^2(t,k) &= \mathrm{Var}(k-T \,|\, t \le T \le k) = \mathrm{Var}(T \,|\, t \le T \le k) , \end{aligned}$$

respectively, where $(t,k) \in D^*$. The function $\tilde{\mu}_d(t,k)$ can be rewritten as follow,

$$\begin{aligned} \tilde{\mu}_d(t,k) &= E(k-T \,|\, t \le T \le k) \\ &= \frac{(t-k)\,F(t-1) + \sum_{i=t}^{k-1} F(i)}{F(k)-F(t-1)} . \end{aligned} \tag{3.3}$$

One can easily obtain that the doubly truncated reversed residual expected value can characterize the general failure rate $h_2(t,k)$ via the relation,

$$h_2(t,k) = 1 - \frac{\tilde{\mu}_d(t,k)}{1+\tilde{\mu}_d(t,k-1)} , \qquad (t,k) \in D^* . \tag{3.4}$$

Khorashadizadeh *et al.* (2012) have shown that if T be discrete random variable with support $\{0,1,2,...,m\}$ (m can be finite or infinite), then for a known t, $F(\cdot)$ can be uniquely recovered by $\tilde{\mu}_d(t,k)$ as follows:

$$F(k) = A_k + F(t-1)\,[1-A_k] , \tag{3.5}$$

where $A_k = \prod_{i=k+1}^{m} \frac{\tilde{\mu}_d(t,i)}{1+\tilde{\mu}_d(t,i-1)}$ and $F(t-1) = \frac{A_{-1}}{A_{-1}-1}$.

The monotonic ageing classes of distributions, *IdRRM* and *IdRRV* in discrete cases can be defined similar to Definition 2.1. Based on the discrete random variable $T_k^* = (k-T \,|\, T<k)$, Goliforushani and Asadi (2008) showed that the class of decreasing reversed residual expected value is empty.

Theorem 3.1. *There is no non-degenerate discrete distribution that has DdRRM or DdRRV property.*

Proof: On contrary, suppose that $\tilde{\mu}_d(t,k)$ is decreasing in k, then for any fixed t, $\tilde{\mu}_d(t,k+1) \le \tilde{\mu}_d(t,k) \le \tilde{\mu}_d(t,t) = 0$, which is contradict the fact that $\tilde{\mu}_d(t,k) \ge 0$. Similar prove can be done for *DdRRV* property. □

One can obtain that

$$\begin{aligned} \tilde{\sigma}_d^2(t,k) &= \frac{(k+t)\,(k-t+1)\,F(t-1) - 2\sum_{i=t}^{k} i\,F(i-1)}{F(k)-F(t-1)} \\ &\quad + (2k+1)\,\tilde{\mu}_d(t,k) - \tilde{\mu}_d^2(t,k) . \end{aligned} \tag{3.6}$$

In the next theorem, we obtain an upper bound for $\tilde{\sigma}_d^2(t,k)$, when $\tilde{\mu}_d(t,k)$ is increasing in t, which is not the same as that obtained in Theorem 2.2 for continuous case.

Theorem 3.2. *If the non-negative discrete random variable T, has $IdRRM$ property, then,*

$$\tilde{\sigma}_d^2(t,k) < \tilde{\mu}_d(t,k)\,[1+\tilde{\mu}_d(t,k)]\;, \qquad (t,k)\in D^*\;. \tag{3.7}$$

Proof: According to (3.3), we have

$$\begin{aligned} 2\sum_{i=t}^{k}\big[F(i)-F(t-1)\big]\,\tilde{\mu}_d(t,i) \;&=\; (k+t)(k-t+1)F(t-1) - 2\sum_{j=t}^{k} jF(j-1) \\ &\quad + (2k+2)\left[(t-k)\,F(t-1)+\sum_{j=t}^{k-1}F(j)\right]. \end{aligned} \tag{3.8}$$

Thus, dividing the both sides of (3.8) by $F(k)-F(t-1)$ and making use of (3.6), implies

$$\begin{aligned} \tilde{\sigma}_d^2(t,k)-\tilde{\mu}_d^2(t,k) \;&=\; \frac{2}{F(k)-F(t-1)}\sum_{i=t}^{k-1}\big[F(i)-F(t-1)\big]\big[\tilde{\mu}_d(i,k)-\tilde{\mu}_d(t,k)\big] \\ &\quad + \tilde{\mu}_d(t,k)\left[\frac{F(k)+F(t-1)}{F(k)-F(t-1)}\right] \\ &< \tilde{\mu}_d(t,k)\;, \end{aligned}$$

since $\tilde{\mu}_d(t,k)$ is increasing with respect k. Hence the required result is obtained. □

The connection between $IdRRV$ and other classes of distributions are also discussed for discrete case in the following theorem.

Theorem 3.3. *In discrete lifetime distributions, the $IdRRM$ property implies the $IdRRV$ property.*

Proof: Using (3.6), we have

$$\begin{aligned} \tilde{\sigma}_d^2(t,k)-\tilde{\sigma}_d^2(t,k-1) \;&=\; \frac{(k+t)\,(k-t+1)\,F(t-1)-2\sum_{i=t}^{k} i\,F(i-1)}{F(k)-F(t-1)} \\ &\quad + \;(2k+1)\,\tilde{\mu}_d(t,k)-\tilde{\mu}_d^2(t,k) \\ &\quad - \;\frac{(k+t-1)\,(k-t)\,F(t-1)-2\sum_{i=t}^{k-1} i\,F(i-1)}{F(k-1)-F(t-1)} \\ &\quad - \;(2k-1)\,\tilde{\mu}_d(t,k-1)+\tilde{\mu}_d^2(t,k-1)\;. \end{aligned} \tag{3.9}$$

On the other hand, one can see that

$$\frac{(k+t)\,(k-t+1)\,F(t-1) - 2\sum_{i=t}^{k} i\,F(i-1)}{F(k) - F(t-1)} -$$

$$- \frac{(k+t-1)\,(k-t)\,F(t-1) - 2\sum_{i=t}^{k-1} i\,F(i-1)}{F(k-1) - F(t-1)} =$$

$$= \Big[(2k-1)\,\tilde{\mu}_d(t,k-1) - \tilde{\mu}_d^2(t,k-1) - \tilde{\sigma}_d^2(t,k-1) + 2k\Big]\,h_2(t,k) - 2k$$

and also

$$\tilde{\mu}_d(t,k) - \tilde{\mu}_d(t,k-1) \;=\; 1 - h_2(t,k)\left[\tilde{\mu}_d(t,k-1) + 1\right] .$$

So, by using these two relations and summarizing the equations, we can write the Eq. (3.9) as

$$\text{(3.10)} \qquad \tilde{\sigma}_d^2(t,k) - \tilde{\sigma}_d^2(t,k-1) \;=\; h_2(t,k)\Big[\tilde{\mu}_d(t,k)\,\tilde{\mu}_d(t,k-1) + \tilde{\mu}_d(t,k) - \tilde{\sigma}_d^2(t,k-1)\Big] .$$

Since, $\tilde{\mu}_d(t,k)$ is increasing in k,

$$\tilde{\sigma}_d^2(t,k) - \tilde{\sigma}_d^2(t,k-1) \;\geq\; h_2(t,k)\Big[\tilde{\mu}_d^2(t,k-1) + \tilde{\mu}_d(t,k-1) - \tilde{\sigma}_d^2(t,k-1)\Big] ,$$

so on using Theorem 3.2, we get the required results. □

The converse of the Theorem 3.3 is not true. The following counterexample shows that *IdRRV* property dose not imply the *IdRRM* property.

Example 3.1. Let T be a discrete random variable with cdf,

k	0	1	2	3	4	5
$F(k)$	0.0625	0.1046	0.1901	0.5561	0.875	1

One can see that in this distribution, $\tilde{\sigma}_d^2(0,k)$ is increasing in k, but $\tilde{\mu}_d(0,k)$ is not monotone.

We consider the discrete doubly truncated reversed residual coefficient of variation as

$$\tilde{\gamma}_d(t,k) \;=\; \frac{\tilde{\sigma}_d(t,k)}{\tilde{\mu}_d(t,k)} .$$

Another characterizations for the *IdRRV* and *IdRRM* classes of distributions based on $\tilde{\gamma}_d(t,k)$ are obtained in the next theorem.

Theorem 3.4. *For non-negative discrete random variable T, we have*

(i) *T has IdRRV property, if and only if,*

$$\tilde{\gamma}_d^2(t,k) \;\le\; \frac{\tilde{\mu}_d(t,k+1)}{\tilde{\mu}_d(t,k)} \left[1 + \frac{1}{\tilde{\mu}_d(t,k)}\right].$$

(ii) *T has IdRRM property, if and only if,*

$$\tilde{\gamma}_d^2(t,k) \;\le\; 1 + \frac{1}{\tilde{\mu}_d(t,k)} \;.$$

Proof: The statement (i) can be proved by using Eq. (3.10) and the statement (ii) can be proved by using Theorem 3.2. □

In the next theorem, we present a characterization via $\tilde{\sigma}_d^2(t,k)$ which is not quite similar to Theorem 2.4 in continuous case.

Theorem 3.5. *$\tilde{\sigma}_d^2(t,k)$ is increasing in k, if and only if,*

$$\frac{\tilde{\sigma}_d^2(t,k-1)}{\tilde{\mu}_d(t,k)\,\tilde{\mu}_d^+(t,k)} \;\le 1\;,$$

where $\tilde{\mu}_d^+(t,k) = E(k-T \,|\, t \le T < k)$.

Proof: Using (3.10) and $\tilde{\mu}_d^+(t,k) = \tilde{\mu}_d(t,k-1)+1$ the required result is obtained. □

4. CHARACTERIZATIONS OF SOME DISCRETE LIFETIME DISTRIBUTIONS

In this section, we characterize discrete distributions based on the doubly truncated random variables. Nair and Sudheesh (2008) and Sudheesh and Nair (2010) have presented some characterization results with their applications for discrete distributions based on one way truncated random variable. In the following theorems, we extend their results for doubly truncated random variables, which are more general and applicable. Since sometimes, the available information is in the specific interval period of times.

Let $c(\cdot)$ be any real valued function, so that for any $(t_1,t_2) \in D^*$,

$$m_c(t_1,t_2) \;=\; E\big(c(T)\,|\,t_1 \le T \le t_2\big) \;=\; \frac{\sum_{i=t_1}^{t_2} c(i)\,p(i)}{F(t_2)-F(t_1-1)} \tag{4.1}$$

is the conditional expected value of doubly (interval) truncated random variable.

Theorem 4.1. *Let T be a non-negative discrete random variable with pmf, $p(t)$, and cdf, $F(t)$. Also suppose that for any real valued function $c(\cdot)$, $\mu = E(c(T))$ and $\sigma^2 = \mathrm{Var}(c(T))$, then for $(t_1, t_2) \in D^*$, T follows the family of distributions satisfying*

$$\frac{p(t+1)}{p(t)} = \frac{\sigma\, g(t)}{\sigma\, g(t+1) - \mu + c(t+1)}\,, \qquad t = 0, 1, 2, \ldots\,, \tag{4.2}$$

with $\sigma\, g(t) = \sum_{i=0}^{t} \frac{p(i)}{p(t)}\,[\mu - c(i)]$, *if and only if,*

$$m_c(t_1, t_2) = \mu + \sigma\, g(t_1 - 1)\, \frac{h_1(t_1 - 1, t_2)}{1 - h_1(t_1 - 1, t_2)} - \sigma\, g(t_2)\, h_2(t_1, t_2)\,. \tag{4.3}$$

Proof: Suppose (4.2) holds, then,

$$c(t)\, p(t) = \mu p(t) + \sigma\, p(t-1)\, g(t-1) - \sigma\, p(t)\, g(t)\,. \tag{4.4}$$

Summation the both side of (4.4) from t_1 to t_2 leads to

$$\sum_{i=t_1}^{t_2} \bigl[c(i) - \mu\bigr]\, p(i) = \sigma \bigl[p(t_1 - 1)\, g(t_1 - 1) - p(t_2)\, g(t_2)\bigr]\,. \tag{4.5}$$

Dividing the both sides of (4.5) by $F(t_2) - F(t_1 - 1)$ and using (3.1), (3.2) and (4.1), we get (4.3) and *vise versa*. □

Remark 4.1. It can be seen that, in special cases, when $t_1 \to 0$ or $t_2 \to \infty$, Theorem 4.1 is identical with that of Nair and Sudheesh (2008) and Sudheesh and Nair (2010).

In the next theorem, we characterize the family of the form (4.2) based on doubly truncated conditional covariance and expected value.

Theorem 4.2. *The distribution function of the non-negative discrete random variable T, belongs to the family of the form (4.2), if and only if, for all non-negative integer values $(t_1, t_2) \in D^*$,*

$$\begin{aligned} \mathrm{Cov}\bigl(s(T), c(T) \,|\, t_1 \le T \le t_2\bigr) &= \sigma\, E\bigl(\Delta s(T).g(T) \,|\, t_1 \le T \le t_2\bigr) \\ &\quad + \bigl[\mu - m_c(t_1, t_2)\bigr]\bigl[m_s(t_1, t_2) - s(t_2 + 1)\bigr] \\ &\quad - \sigma\, g(t_1 - 1)\, \frac{h_1(t_1 - 1, t_2)}{1 - h_1(t_1 - 1, t_2)} \bigl[s(t_2 + 1) - s(t_1)\bigr]\,, \end{aligned} \tag{4.6}$$

where $c(\cdot)$ and $s(\cdot)$ are any real valued functions such that $E(s^2(T)) < \infty$, $E(\Delta s(T)\,.\,g(T)) < \infty$, $\Delta s(T) \neq 0$ and $m_s(t_1, t_2) = E(s(T) \,|\, t_1 \le T \le t_2)$.

Proof: First, we know that

$$(4.7) \qquad E\big(s(T)\,(c(T)-\mu)\,|\,t_1 \le T \le t_2\big) \;=$$

$$= \frac{1}{F(t_2)-F(t_1-1)} \sum_{i=t_1}^{t_2} s(i)\,\big[c(i)-\mu\big]\,p(i)$$

$$= \frac{1}{F(t_2)-F(t_1-1)} \sum_{j=0}^{t_2-t_1} s(t_1+j) \left[\sum_{i=t_1+j}^{t_2} \big(c(i)-\mu\big)p(i) - \sum_{i=t_1+j+1}^{t_2} \big(c(i)-\mu\big)\,p(i)\right]$$

$$= \frac{1}{F(t_2)-F(t_1-1)} \left[\sum_{j=t_1}^{t_2} \Delta s(j) \sum_{i=j+1}^{t_2} \big(c(i)-\mu\big)\,p(i)\right] + s(t_1)\,\big(m_c(t_1,t_2)-\mu\big)\,.$$

Now, suppose that T has a distribution of form (4.2), hence, on using

$$\sum_{i=j+1}^{t_2} \big(c(i)-\mu\big)\,p(i) \;=\; \sigma\big[p(j)\,g(j) - p(t_2)\,g(t_2)\big]\,,$$

we have

$$(4.8) \qquad E\big(s(T)\,(c(T)-\mu)\,|\,t_1 \le T \le t_2\big) \;=$$

$$= \frac{1}{F(t_2)-F(t_1-1)} \left[\sigma \sum_{j=t_1}^{t_2} \Delta s(j)\,g(j)\,p(j) - \sigma\,g(t_2)\,p(t_2) \sum_{j=t_1}^{t_2} \Delta s(j)\right]$$

$$+\; s(t_1)\,\big(m_c(t_1,t_2)-\mu\big)$$

$$= \sigma\,E\big(\Delta s(T)\,.\,g(T)\,|\,t_1 \le T \le t_2\big) \;+\; \sigma\,s(t_1)\,g(t_1-1)\,\frac{h_1(t_1-1,t_2)}{1-h_1(t_1-1,t_2)}$$

$$+\; s(t_2+1)\left[m_c(t_1,t_2) - \mu - \sigma\,g(t_1-1)\,\frac{h_1(t_1-1,t_2)}{1-h_1(t_1-1,t_2)}\right],$$

or

$$E\big(s(T)\,.\,c(T)\,|\,t_1 \le T \le t_2\big) \;=$$

$$= \sigma\,E\big(\Delta s(T)\,.\,g(T)\,|\,t_1 \le T \le t_2\big) \;+\; s(t_2+1)\,(m_c(t_1,t_2)-\mu)$$

$$-\; \sigma\,g(t_1-1)\,\frac{h_1(t_1-1,t_2)}{1-h_1(t_1-1,t_2)}\,\big[s(t_2+1)-s(t_1)\big] \;+\; \mu\,m_s(t_1,t_2)\,,$$

which easily leads to (4.6).

Conversely, let (4.6) is true, then, comparing (4.7) and (4.8) implies

$$(4.9) \qquad \sum_{j=t_1}^{t_2} \Delta s(j) \sum_{i=j+1}^{t_2} \big(c(i)-\mu\big)\,p(i) \;=$$

$$= \sigma \sum_{j=t_1}^{t_2} \Delta s(j)\,g(j)\,p(j) - \sigma\,g(t_2)\,p(t_2) \sum_{j=t_1}^{t_2} \Delta s(j)\,.$$

Changing t_1 to t_1-1 and subtracting from (4.9) leads to (4.5), which is equivalent to the distribution with the form (4.2). $\square$

Remark 4.2. In special case of Theorem 4.2, when $s(T) = c(T) = T$, we have

$$\begin{aligned} \tilde{\sigma}_d^2(t_1,t_2) &= \mathrm{Var}\big(T \,|\, t_1 \le T \le t_2\big) \\ &= \sigma^* E\big(g(T) \,|\, t_1 \le T \le t_2\big) + \big[\mu^* - m(t_1,t_2)\big]\,\big[m(t_1,t_2) - t_2 - 1\big] \\ &\quad - \sigma^* g(t_1-1)\,\frac{h_1(t_1-1,t_2)}{1-h_1(t_1-1,t_2)}\,[t_2 - t_1 + 1]\,, \end{aligned} \tag{4.10}$$

where $m(t_1,t_2) = E(T \,|\, t_1 \le T \le t_2)$ is the doubly truncated expected time to failure function and $\mu^* = E(T)$ and $\sigma^{2*} = \mathrm{Var}(T)$.

In the table in the following page, we illustrate the results of Remark 4.2 in some distributions.

Remark 4.3. It should be noted that similar results and definitions in discrete case, can be verified by using the doubly truncated reversed random variables ${}_{t^+}T_k = (k - T \,|\, t < T \le k)$, ${}_{t}T_{k^-} = (k - T \,|\, t \le T < k)$ or ${}_{t^+}T_{k^-} = (k - T \,|\, t < T < k)$.

5. SUMMARY AND CONCLUSIONS

In this paper, we obtain some reliability properties of the reversed residual lifetime via doubly truncation. Also, their similarities and differences are compared in both discrete and continuous lifetime distributions and the following partial chain is obtained.

$$h_2(a,b) \text{ is decreasing in } b \implies IdRRM \implies IdRRV\,.$$

Also, some characterization results are obtained in discrete distributions via conditional covariance and variance.

ACKNOWLEDGMENTS

The authors are grateful to the associate editor and referees for their valuable comments which improved the article. Also, the support of Ordered and Spatial Data Center of Excellence of Ferdowsi University of Mashhad is acknowledged.

Distribution	$p(t)$	$\sigma^* g(t)$	$\mathrm{Var}\big(T \mid t_1 \le T \le t_2\big)$
Binomial	$\binom{n}{t} p^t (1-p)^{n-t}$ $t = 0, 1, \ldots, n$ $0 \le p \le 1$	$(n-t)\,p$	$p\left[m(t_1,t_2)\,(n-1) - nt_2 - \frac{h_1(t_1-1,t_2)}{1-h_1(t_1-1,t_2)}\,(t_2-t_1+1)\,(n-t_1+1)\right] - $ $-\,m(t_1,t_2)\,\big(m(t_1,t_2)-t_2-1\big)$
Poisson	$\frac{e^{-\theta}\,\theta^t}{t!}$ $t = 0, 1, \ldots$ $\theta \ge 0$	θ	$\theta\left[m(t_1,t_2) - \frac{t_2}{1-h_1(t_1-1,t_2)} + (t_1-1)\,\frac{h_1(t_1-1,t_2)}{1-h_1(t_1-1,t_2)}\right] - $ $-\,m(t_1,t_2)(m(t_1,t_2)-t_2-1)$
Neg. binomial	$\binom{k+t-1}{k-1}\,\theta^k\,(1-\theta)^t$ $t = 0, 1, 2, \ldots$ $0 \le \theta \le 1$	$\frac{1-\theta}{\theta}\,(t+k)$	$\frac{1}{\theta}\left[(t_1+k-1)\,(t_2-t_1+1)\,(\theta-1)\,\frac{h_1(t_1-1,t_2)}{1-h_1(t_1-1,t_2)} + \big(m(t_1,t_2)-t_2\big)\,k + m(t_1,t_2)\right] - $ $-\,(m(t_1,t_2)+k)(m(t_1,t_2)-t_2)$
Geometric	$\theta\,(1-\theta)^t$ $t = 0, 1, 2, \ldots$ $0 \le \theta \le 1$	$\frac{1-\theta}{\theta}\,(t+1)$	$\frac{t_1(t_2-t_1+1)}{1-(1-\theta)^{t_2-t_1+1}} + \frac{2\,m(t_1,t_2)-t_2}{\theta}\,\big(m(t_1,t_2)+1\big)\,\big(m(t_1,t_2)-t_2\big)$

REFERENCES

[1] Ahmad, A. A. (2001). Moments of order statistics from doubly truncated continuous distributions and characterizations, *Statist.*, **35**(4), 479–494.

[2] Bairamov, I. and Gebizlioglu, O. (2005). *On the characterizations of distributions through the properties of conditional expectations of order statistics*, In "Advances on Models, Characterizations and Applications" (N. Balakrishnan, I. Bairamov and O.L. Gebizlioglu, Eds.), Boca Raton, FL: Chapman and Hall / CRC Press.

[3] Betensky, R. A. and Martin, E. C. (2003). Commentary: Failure-rate functions for doubly truncated random variables, *IEEE Trans. on Reliab.*, **52**(1), 7–8.

[4] Golifroushani, S. and Asadi, M. (2008). On the discrete mean past lifetime, *Metrika*, **68**, 209–217.

[5] Khorashadizadeh, M.; Rezaei Roknabadi, A. H. and Mohtashami Borzadaran, G. R. (2012). Characterizations of lifetime distributions based on doubly truncated mean residual life and mean past to failure, *Commun. Statist. Theor. Meth.*, **41**(6), 1105–1115.

[6] Nair, U. N. and Sudheesh, K. K. (2008). Some results on lower variance bounds useful in reliability modeling and estimation, *Ann. Inst. Stat. Math.*, **60**, 591–603.

[7] Nanda, A. K.; Singh, H.; Misra, N. and Paul, P. (2003). Reliability properties of reversed residual lifetime, *Commun. Statist. Theor. Meth.*, **32**(10), 2031–2042.

[8] Navarro, J. and Ruiz, J. M. (1996). Failure-Rate Function for Doubly-Truncated Random Variables, *IEEE Trans. Reliab.*, **45**(4), 685–690.

[9] Navarro, J. and Ruiz, J. M. (2004). Characterization from relationships between failure rate functions and conditional moments, *Commun. Statist. Theor. Meth.*, **33**(12), 3159–3171.

[10] Navarro, J.; Ruiz, J. M. and Zoroa, N. (1998). A unified approach to characterization problems using conditional expectations, *J. Statist. Plann. Infer.*, **69**, 193–207.

[11] Poursaeed, M. H. and Nematollahi, A. R. (2008). On the mean past and the mean residual life under double monitoring, *Commun. Statist. Theor. Meth.*, **37**(7), 1119–1133.

[12] Ruiz, J. M. and Navarro, J. (1995). Characterization of discrete distributions using expectated values, *Statistical Papers*, **36**, 237–252.

[13] Ruiz, J. M. and Navarro, J. (1996). Characterization based on conditional expectations of the doubled truncated distribution, *Ann. Ins. Statist. Math.*, **48**(3), 563–572.

[14] Sankaran, P. G. and Sunoj, S. M. (2004). Identification of models using failure rate and mean residual life of doubly truncated random variables, *Statist. Pap.*, **45**, 97–109.

[15] Su, J. C. and Huang, W. J. (2000). Characterizations based on conditional expectations, *Statist. Pap.*, **41**, 423–435.

[16] Sudheesh, K. K. and Nair, N. U. (2010). Characterization of discrete distributions by conditional variance, *Metron*, **LXVIII**(1), 77–85.

[17] Sunoj, S. M.; Sankaran, P. G. and Maya, S. S. (2009). Characterizations of life distributions using conditional expectations of doubly (interval) truncated random variables, *Commun. Statist. Theor. Meth.*, **38**(9), 1441–1452.

REVSTAT – Statistical Journal
Volume 15, Number 1, January 2017, 107–139

INTRIGUING PROPERTIES OF EXTREME GEOMETRIC QUANTILES

Authors: Stéphane Girard
– Team Mistis, Inria Grenoble Rhône-Alpes & LJK, France
Stephane.Girard@inria.fr

Gilles Stupfler
– GREQAM UMR 7316, CNRS, EHESS, Centrale Marseille, Aix Marseille Université, France
gilles.stupfler@univ-amu.fr

Received: January 2015 Revised: May 2015 Accepted: May 2015

Abstract:

- Central properties of geometric quantiles have been well-established in the recent statistical literature. In this study, we try to get a grasp of how extreme geometric quantiles behave. Their asymptotics are provided, both in direction and magnitude, under suitable moment conditions, when the norm of the associated index vector tends to one. Some intriguing properties are highlighted: in particular, it appears that if a random vector has a finite covariance matrix, then the magnitude of its extreme geometric quantiles grows at a fixed rate. We take profit of these results by defining a parametric estimator of extreme geometric quantiles of such a random vector. The consistency and asymptotic normality of the estimator are established, and contrasted with what can be obtained for univariate quantiles. Our results are illustrated on both simulated and real data sets. As a conclusion, we deduce from our observations some warnings which we believe should be known by practitioners who would like to use such a notion of multivariate quantile to detect outliers or analyze extremes of a random vector.

Key-Words:

- *extreme quantile; geometric quantile; consistency; asymptotic normality.*

AMS Subject Classification:

- 62H05, 62G20, 62G32.

1. INTRODUCTION

Let X be a random vector in $\mathbb{R}^d$. Up to now, several definitions of multivariate quantiles of X have been proposed in the statistical literature. We refer to [25] for a review of various possibilities for this notion. Here, we focus on the notion of "spatial" or "geometric" quantiles, introduced by [14], which generalises the characterisation of a univariate quantile shown in [22]. For a given vector u belonging to the unit open ball B^d of $\mathbb{R}^d$, where $d \geq 2$, a geometric quantile with index vector u is any solution of the optimisation problem defined by

$$\arg\min_{q \in \mathbb{R}^d} \mathbb{E}\big(\|X - q\| - \|X\|\big) - \langle u, q \rangle \,, \tag{1.1}$$

where $\langle \cdot, \cdot \rangle$ is the usual scalar product on $\mathbb{R}^d$ and $\|\cdot\|$ is the associated Euclidean norm. Note that $q(u) \in \mathbb{R}^d$ possesses both a direction and magnitude. It can be seen that geometric quantiles are in fact special cases of M-quantiles introduced by [3] which were further analysed by [23]. Besides, such quantiles have various strong properties. First, the quantile with index vector $u \in B^d$ is unique whenever the distribution of X is not concentrated on a single straight line in $\mathbb{R}^d$ (see [14] or Theorem 2.17 in [21]). Second, although they are not fully affine equivariant, they are equivariant under any orthogonal transformation [14]. Third, geometric quantiles characterise the associated distribution. Namely, if two random variables X and Y yield the same quantile function q, then X and Y have the same distribution [23]. Finally, for $u = 0$, the well-known L^2-geometric median is obtained, which is the simplest example of a "central" quantile [28]. We point out that one may compute an estimation of the geometric median in an efficient way, see [8].

These properties make geometric quantiles reasonable candidates when trying to define multivariate quantiles, which is why their estimation was studied in several papers. We refer for instance to [14], who established a Bahadur expansion for the estimator of geometric quantiles obtained by solving the sample counterpart of problem (1.1). [10] then introduced a transformation–retransformation procedure to obtain affine equivariant estimates of multivariate quantiles. This notion was extended to a multiresponse linear model by [11]. Recently, [16] defined a multivariate quantile–quantile plot using geometric quantiles. Conditional geometric quantiles can also be defined by substituting a conditional expectation to the expectation in (1.1). We refer to [6] for the estimation of the conditional geometric median and to [15] for the estimation of an arbitrary conditional geometric quantile. The estimation of a conditional median when there is an infinite-dimensional covariate is considered in [13].

Let us note though that the previous papers focus on central properties of geometric quantiles and of their sample versions. While some of them label geometric quantiles as "extreme" when $\|u\|$ is close to 1 ([14, 15]) and use it in real applications (see *e.g.* [12] for an application to outlier detection), the specific properties of these extreme geometric quantiles have not been investigated yet. In this study, we provide the asymptotics of the direction and magnitude of the extreme geometric quantile $q(u)$ when $\|u\| \to 1$, under suitable moment conditions. There are well-known analogue results for univariate extreme quantiles in the right tail of a distribution, see *e.g.* [18]. A particular corollary of our results is that the magnitude of the extreme geometric quantiles of a random vector X having a finite covariance matrix grows at a fixed rate. Moreover, in this case, the magnitude of the extreme geometric quantiles is asymptotically characterised by the covariance matrix of X. This is an intriguing property, which opens the door to a parametric estimation of extreme quantiles whose asymptotic properties are studied in this work.

The outline of the paper is as follows. Asymptotic properties of geometric quantiles are stated in Section 2. An illustrative application to the estimation of extreme geometric quantiles is given in Section 3. Some examples and numerical illustrations of our results, including a study of a real data set, are presented in Section 4. Section 5 offers a couple of concluding remarks, in which some warnings are given to practitioners who would like to use such geometric quantiles to detect outliers or analyze extremes of a random vector. Proofs are deferred to Section 6.

2. ASYMPTOTIC BEHAVIOUR OF EXTREME GEOMETRIC QUANTILES

From now on, we assume that the distribution of X is not concentrated on a single straight line in $\mathbb{R}^d$ and non-atomic. [14] proved that, in this context, the solution $q(u)$ of (1.1), namely the geometric quantile with index vector u, exists and is unique for every $u \in B^d$. Let $\psi\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be defined as $\psi(u,q) = \mathbb{E}(\|X-q\| - \|X\|) - \langle u, q\rangle$ and assume further that $t/\|t\| = 0$ if $t = 0$. If $u \in \mathbb{R}^d$ is such that there is a solution $q(u) \in \mathbb{R}^d$ to problem (1.1), then the gradient of $q \mapsto \psi(u,q)$ must be zero at $q(u)$, that is

$$u + \mathbb{E}\left(\frac{X - q(u)}{\|X - q(u)\|}\right) = 0\,. \tag{2.1}$$

This condition immediately entails that if $u \in \mathbb{R}^d$ is such that problem (1.1) has a solution $q(u)$, then $\|u\| \le 1$. In fact, we can prove a stronger result:

Proposition 2.1. *The optimisation problem (1.1) has a solution if and only if $u \in B^d$.*

Moreover, remarking that the function $\psi(u,\cdot)$ is strictly convex, [14] proved the following characterisation of a geometric quantile: for every $u \in B^d$, $q(u)$ is the solution of problem (1.1) if and only if it satisfies equation (2.1). In particular, this entails that the function $G\colon \mathbb{R}^d \to B^d$ defined by

$$\forall q \in \mathbb{R}^d\,, \qquad G(q) = -\mathbb{E}\left(\frac{X-q}{\|X-q\|}\right)$$

is a continuous bijection. Proposition 2.6(iii) in [23] shows that the inverse of the function G, *i.e.* the geometric quantile function $u \mapsto q(u)$, is also continuous on B^d.

In most cases however, computing explicitly the function G is a hopeless task, which makes it impossible to obtain a closed-form expression for the geometric quantile function. It is thus of interest to prove general results about the geometric quantile $q(u)$, especially regarding its direction and magnitude. Our first main result focuses on the special case of spherically symmetric distributions.

Proposition 2.2. *If X has a spherically symmetric distribution then:*

- **(i)** *The map $u \mapsto q(u)$ commutes with every linear isometry of $\mathbb{R}^d$. Especially, the norm of a geometric quantile $q(u)$ only depends on the norm of u.*
- **(ii)** *For all $u \in B^d$, the geometric quantile $q(u)$ has direction u if $u \neq 0$ and $q(0) = 0$ otherwise.*
- **(iii)** *The function $\|u\| \mapsto \|q(u)\|$ is a continuous strictly increasing function on $[0,1)$.*
- **(iv)** *It holds that $\|q(u)\| \to \infty$ as $\|u\| \to 1$.*

Although Proposition 2.2(i,iii) cannot be expected to hold true for a random variable which is not spherically symmetric, one may wonder if (ii,iv), namely that a geometric quantile shares the direction of its index vector and that the norm of the geometric quantile function tends to infinity on the unit sphere, can be extended to the general case. The next result, which examines the behaviour of the geometric quantile function near the boundary of the open ball B^d, provides an answer to this question.

Theorem 2.1. *Let S^{d-1} be the unit sphere of $\mathbb{R}^d$.*

- **(i)** *It holds that $\|q(v)\| \to \infty$ as $\|v\| \to 1$.*
- **(ii)** *Moreover, if $v \to u$ with $u \in S^{d-1}$ and $v \in B^d$ then $q(v)/\|q(v)\| \to u$.*

Theorem 2.1 shows two properties of geometric quantiles: first, the norm of the geometric quantile $q(v)$ with index vector v diverges to infinity as $\|v\| \uparrow 1$. In other words, Proposition 2.2(iv) still holds for any distribution. This is a rather intriguing property of geometric quantiles, since it holds even if the distribution of X has a compact support (for instance, when X is uniformly distributed on a square). A related point is the fact that sample geometric quantiles do not necessarily lie within the convex hull of the sample, see [4] for a counter-example. Second, if $v \to u \in S^{d-1}$, then the geometric quantile $q(v)$ has asymptotic direction u. Proposition 2.2(ii) thus remains true asymptotically for any distribution.

It is possible to specify the convergences obtained in Theorem 2.1 under moment assumptions. Theorem 2.2 provides a first-order expansion of both the direction and the magnitude of an extreme geometric quantile $q(\alpha u)$ in the direction u, where u is a unit vector and α tends to 1.

Theorem 2.2. *Let $u \in S^{d-1}$.*

(i) *If $\mathbb{E}\|X\| < \infty$ then $q(\alpha u) - \big\{\|q(\alpha u)\| u + \mathbb{E}(X - \langle X, u\rangle u)\big\} \to 0$ as $\alpha \uparrow 1$.*

(ii) *If $\mathbb{E}\|X\|^2 < \infty$ and Σ denotes the covariance matrix of X then*

$$\|q(\alpha u)\|^2 (1-\alpha) \;\to\; \frac{1}{2}\big(\operatorname{tr}\Sigma - u'\Sigma u\big) > 0 \qquad \text{as} \quad \alpha \uparrow 1\,.$$

Let us note that the integrability conditions of Theorem 2.2 exclude any random vector $\|X\|$ whose distribution possesses a right tail which is too heavy. For instance, condition $\mathbb{E}\|X\| < \infty$ in (i) excludes the multivariate Student distribution with less than one degree of freedom, while condition $\mathbb{E}\|X\|^2 < \infty$ in (ii) excludes the multivariate Student distribution with less than two degrees of freedom.

Consequence 1. It appears that, if X has a finite covariance matrix Σ, then the magnitude of an extreme geometric quantile is determined (in the asymptotic sense) by Σ. In other words, since the asymptotic direction of an extreme geometric quantile in the direction u is exactly u by Theorem 2.1, it follows that the extreme geometric quantiles of two probability distributions which admit the same finite covariance matrix are asymptotically equivalent. This phenomenon is illustrated on simulated data in Section 4 below. This is surprising from the extreme value perspective: one could expect the behaviour of extreme geometric quantiles not to be driven by a central parameter such as the covariance matrix, as happens in the univariate context where the value of an extreme quantile depends on the tail heaviness of the probability density function of X.

Consequence 2. The map $\lambda \mapsto \|q((1-\lambda^{-1})u)\|$ is a regularly varying function with index 1/2 (see Bingham *et al.*, 1987) and therefore:

$$\frac{\|q(\beta u)\|}{\|q(\alpha u)\|} = \left(\frac{1-\alpha}{1-\beta}\right)^{1/2} \big(1 + \mathrm{o}(1)\big)$$

when $\alpha \to 1$ and $\beta \to 1$. In other words, given an arbitrary extreme geometric quantile, one can deduce the asymptotic behaviour of every other extreme geometric quantile sharing its direction, independently of the distribution.

Again, this is fundamentally different from the univariate case when deducing the value of an extreme quantile from another one then requires the knowledge (or an estimate) of the extreme-value index of the distribution, see [18], Chapter 4. A further, perhaps unexpected, consequence is that our results can actually be used to define a consistent and asymptotically Gaussian estimator of extreme geometric quantiles by using the standard empirical estimator of the covariance matrix of X, see Section 3 below.

Consequence 3. Finally, Theorem 2.2 provides some information on the shape of an extreme quantile contour. It is readily seen that the global maximum of the function $h_1(u) := \operatorname{tr}\Sigma - u'\Sigma u$ on S^{d-1} is reached at a unit eigenvector $u_{\min}$ of Σ associated with its smallest eigenvalue $\lambda_{\min} > 0$. Thus, the norm of an extreme geometric quantile is asymptotically the largest in the direction where the variance is the smallest. Similarly, the global minimum of h_1 is reached at a unit eigenvector $u_{\max}$ of Σ associated with its largest eigenvalue $\lambda_{\max} > 0$. In particular, if f is the probability density function associated with an elliptically contoured distribution [7], the level sets of f coincide with the level sets of the function $h_2(u) := u'\Sigma u$. The global maximum of h_2 is reached at the eigenvector $u_{\max}$ while the global minimum is reached at $u_{\min}$. The extreme geometric quantile is therefore furthest from the origin in the direction where the density level set is closest to the origin, see Section 4 for an illustration on real data. In such a case, the extreme geometric quantile contour plot and the density level plots are in some sense orthogonal (even though they agree when the distribution of X is spherically symmetric). Of course, one should not expect a direct geometric match between quantile contours and density contours, but this phenomenon should be kept in mind when designing outlier detection procedures. In our view, this can be seen as a consequence of the lack of affine-equivariance of geometric quantiles. To tackle this issue, one may apply a transformation–retransformation procedure, see [27]. Such procedures admit sample analogues, see for instance [9, 10], at the possible loss of geometric interpretation, see [26].

3. AN ESTIMATOR OF EXTREME GEOMETRIC QUANTILES

In this paragraph, our focus is to illustrate Consequence 2 of Theorem 2.2 at the sample level. Let $X_1, ..., X_n$ be independent random copies of a random vector X having a finite covariance matrix Σ. It follows from Theorem 2.2 that any extreme geometric quantile $q(\alpha u)$ of X, with $\alpha \uparrow 1$ and $u \in S^{d-1}$ can be approximated by:

$$q_{\mathrm{eq}}(\alpha u) := (1-\alpha)^{-1/2} \left[\frac{1}{2}\left(\operatorname{tr}\Sigma - u'\Sigma u\right)\right]^{1/2} u \,. \tag{3.1}$$

This can be used to define an estimator of the extreme geometric quantiles of X: let $\overline{X}_n = n^{-1}\sum_{k=1}^n X_k$ be the sample mean and

$$\widehat{\Sigma}_n \;=\; \frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X}_n)\,(X_k - \overline{X}_n)'$$

be the empirical estimator of the covariance matrix Σ of X. Let further (α_n) be an increasing sequence of positive real numbers tending to 1. Our estimator $\widehat{q}_n(\alpha_n u)$ of $q(\alpha_n u)$ is then

$$\widehat{q}_n(\alpha_n u) \;=\; (1-\alpha_n)^{-1/2}\left[\frac{1}{2}\left(\operatorname{tr}\widehat{\Sigma}_n - u'\widehat{\Sigma}_n u\right)\right]^{1/2} u\,.$$

The consistency of $\widehat{q}_n(\alpha_n u)$ is examined in the next result.

Theorem 3.1. *Let $u \in S^{d-1}$ and assume that $\alpha_n \uparrow 1$. If $\mathbb{E}\|X\|^2 < \infty$ then*

$$\sqrt{1-\alpha_n}\,\big(\widehat{q}_n(\alpha_n u) - q(\alpha_n u)\big) \;\to\; 0 \qquad \textit{almost surely as} \quad n\to\infty\,.$$

This result actually means that the extreme geometric quantile estimator is relatively consistent in the sense that

$$\frac{\widehat{q}_n(\alpha_n u) - q(\alpha_n u)}{\|q(\alpha_n u)\|} \to 0 \qquad \text{almost surely as} \quad n\to\infty\,,$$

since $\|q(\alpha_n u)\|^{-1} = \mathrm{O}(\sqrt{1-\alpha_n})$, see Theorem 2.2(ii). This normalisation could be expected since the quantity to be estimated diverges in magnitude. Under the additional assumption that X has a finite fourth moment, an asymptotic normality result can be established for this estimator:

Theorem 3.2. *Let $u \in S^{d-1}$ and assume that $\alpha_n \uparrow 1$ is such that $n(1-\alpha_n) \to 0$. If $\mathbb{E}\|X\|^4 < \infty$ then*

$$\sqrt{n(1-\alpha_n)}\,\big(\widehat{q}_n(\alpha_n u) - q(\alpha_n u)\big) \xrightarrow{\;d\;} Z \qquad \textit{as} \quad n\to\infty$$

where Z is a Gaussian centred random vector.

Let us highlight that the covariance matrix of the Gaussian limit in Theorem 3.2 essentially depends on the covariance matrix M of the Gaussian limit of $\sqrt{n}(\widehat{\Sigma}_n - \Sigma)$, see the proof in Section 6. Although M has a complicated expression (see *e.g.* [24]), it can be estimated when $\mathbb{E}\|X\|^4 < \infty$, which makes it possible to construct asymptotic confidence regions for extreme geometric quantiles.

Extreme geometric quantiles can thus be consistently estimated by $\widehat{q}_n(\alpha_n u)$, whatever the "order" α_n, and an asymptotic normality result is obtained when

$\alpha_n \uparrow 1$ *quickly enough.* The proposed estimator is therefore able to extrapolate arbitrarily far from the original sample. This is very different from the univariate case, where the empirical quantile $\widehat{q}_n(\alpha_n) = \inf\{t \in \mathbb{R} \mid \widehat{F}_n(t) \geq \alpha_n\}$, deduced from the empirical cumulative distribution function $\widehat{F}_n$, estimates the true quantile $q(\alpha_n)$ consistently only if α_n converges to 1 *slowly enough.* The extrapolation with faster rates α_n is then handled assuming that the underlying distribution function is heavy-tailed and by using adapted estimators, see *e.g.* [29] and the monograph [18].

4. NUMERICAL ILLUSTRATIONS

4.1. Simulation study

In this section, our main results are illustrated, particularly Theorems 2.2, 3.1 and 3.2 in the bivariate case $d = 2$ to make the display easier. In this framework, $u \in S^1$ can be represented by an angle: $u = u_\theta = (\cos\theta, \sin\theta)$, $\theta \in [0, 2\pi)$. The iso-quantile curves $\mathcal{C}q(\alpha) = \{q(\alpha u_\theta),\ \theta \in [0, 2\pi)\}$ and their estimates $\mathcal{C}\widehat{q}_n(\alpha) = \{\widehat{q}_n(\alpha u_\theta),\ \theta \in [0, 2\pi)\}$ can then be considered in order to get a grasp of the behaviour of extreme quantiles in every direction. The following two distributions are considered for the random vector X:

- The centred Gaussian multivariate distribution $\mathcal{N}(0, v_X, v_Y, v_{XY})$, with probability density function: $\forall x, y \in \mathbb{R}$,

$$f(x,y) = \frac{1}{2\pi\sqrt{\det\Sigma}} \exp\left(-\frac{1}{2}\begin{pmatrix} x \\ y \end{pmatrix}' \Sigma^{-1} \begin{pmatrix} x \\ y \end{pmatrix}\right) \quad \text{with } \Sigma = \begin{pmatrix} v_X & v_{XY} \\ v_{XY} & v_Y \end{pmatrix}.$$

- A double exponential distribution $\mathcal{E}(\lambda_-, \mu_-, \lambda_+, \mu_+)$, with $\lambda_-, \mu_-, \lambda_+, \mu_+ > 0$, whose probability density function is: $\forall x, y \in \mathbb{R}$,

$$f(x,y) = \frac{1}{4}\begin{cases} \lambda_+ \mu_+ e^{-\lambda_+|x| - \mu_+|y|} & \text{if } xy > 0\,, \\ \lambda_- \mu_- e^{-\lambda_-|x| - \mu_-|y|} & \text{if } xy \leq 0\,. \end{cases}$$

  In this case, X is centred and has covariance matrix

$$\Sigma = \begin{pmatrix} \dfrac{1}{\lambda_-^2} + \dfrac{1}{\lambda_+^2} & \dfrac{1}{2}\left[\dfrac{1}{\lambda_+\mu_+} - \dfrac{1}{\lambda_-\mu_-}\right] \\ \dfrac{1}{2}\left[\dfrac{1}{\lambda_+\mu_+} - \dfrac{1}{\lambda_-\mu_-}\right] & \dfrac{1}{\mu_-^2} + \dfrac{1}{\mu_+^2} \end{pmatrix}.$$

Three different sets of parameters were used for each distribution, in order that the related covariance matrices coincide:

- $\mathcal{N}(0,1/2,1/2,0)$ and $\mathcal{E}(2,2,2,2)$ with spherical covariance matrices;
- $\mathcal{N}(0,1/8,3/4,0)$ and $\mathcal{E}(4,2\sqrt{2/3},4,2\sqrt{2/3})$ with diagonal covariance matrices;
- $\mathcal{N}(0,1/2,1/2,1/6)$ and $\mathcal{E}(2\sqrt{3},2\sqrt{3},2\sqrt{3/5},2\sqrt{3/5})$ with full covariance matrices.

In each case, we carry out the following computations:

- For each $\alpha \in \{0.99, 0.995, 0.999\}$, the true quantile curves $\mathcal{C}q(\alpha)$ obtained by solving problem (1.1) numerically, as well as their analogues $\mathcal{C}q_{\text{eq}}(\alpha)$ using approximation (3.1) are computed. The normalised squared approximation error

$$e(\alpha) \;=\; (1-\alpha)\int_0^{2\pi} \|q_{\text{eq}}(\alpha u_\theta) - q(\alpha u_\theta)\|^2 \, d\theta$$

  is then recorded.
- For each value of α, we draw $N = 1000$ replications of an n-sample $(X_1, ..., X_n)$ of independent copies of X, with $n \in \{100, 200, 500\}$. The estimated quantile curves $\mathcal{C}\widehat{q}_n^{(j)}(\alpha)$ corresponding to the j-th replication and the associated normalised squared error

$$E_n^{(j)}(\alpha) \;=\; (1-\alpha)\int_0^{2\pi} \left\|\widehat{q}_n^{(j)}(\alpha u_\theta) - q(\alpha u_\theta)\right\|^2 d\theta$$

  are computed as well as the mean squared error $E_n(\alpha) = N^{-1}\sum_{j=1}^N E_n^{(j)}(\alpha)$.

The true quantile curves, as well as the approximated and the estimated ones are displayed on Figures 1–6 in the case $n = 200$ and $\alpha = 0.995$. The true quantile curves look very similar in Figures 1 and 4, in Figures 2 and 5 and Figures 3 and 6 (in which the words "best", "median" and "worst" are to be understood with respect to the L^2 error). This is in accordance with Theorem 2.2: eventually, extreme geometric quantiles only depend on the covariance matrix of the underlying distribution. Moreover, the approximated quantile curves are close to the true ones in all cases, and the estimated quantile curves are satisfying in all situations with a moderate variability. Similar results were observed for $n = 100, 500$ and $\alpha = 0.99, 0.999$. We do not report the graphs here for the sake of brevity; we do however display the approximation and estimation errors in Table 1. Unsurprisingly, the estimation error $E_n(\alpha)$ decreases as the sample size n increases. Both approximation and estimation errors $e(\alpha)$ and $E_n(\alpha)$ have a stable behaviour with respect to α.

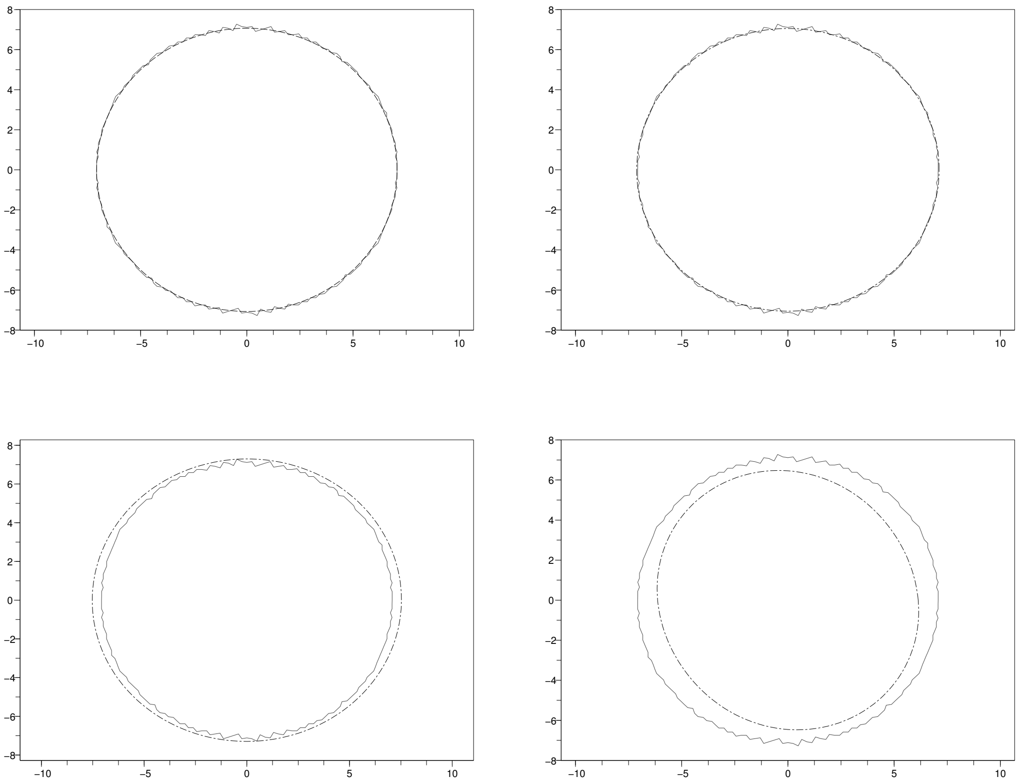


Figure 1: Spherical Gaussian distribution $\mathcal{N}(0, 1/2, 1/2, 0)$ for $\alpha = 0.995$. Top left: comparison between a numerical method and the use of the equivalent (3.1) for the computation of the iso-quantile curve, full line: numerical method, dashed line: asymptotic equivalent. Top right, bottom left and bottom right: best, median and worst estimates of the iso-quantile curve for $n = 200$, full line: numerical method, dashed-dotted line: estimator $\widehat{q}_n$.

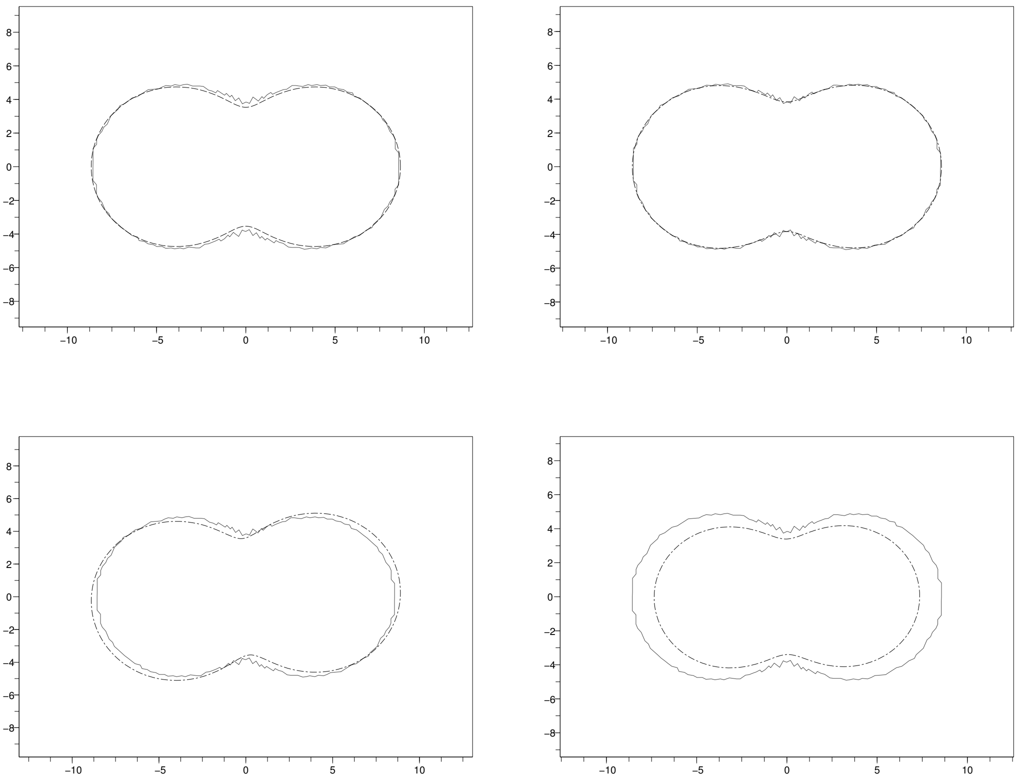


Figure 2: Diagonal Gaussian distribution $\mathcal{N}(0, 1/8, 3/4, 0)$ for $\alpha = 0.995$. Top left: comparison between a numerical method and the use of the equivalent (3.1) for the computation of the iso-quantile curve, full line: numerical method, dashed line: asymptotic equivalent. Top right, bottom left and bottom right: best, median and worst estimates of the iso-quantile curve for $n = 200$, full line: numerical method, dashed-dotted line: estimator $\widehat{q}_n$.

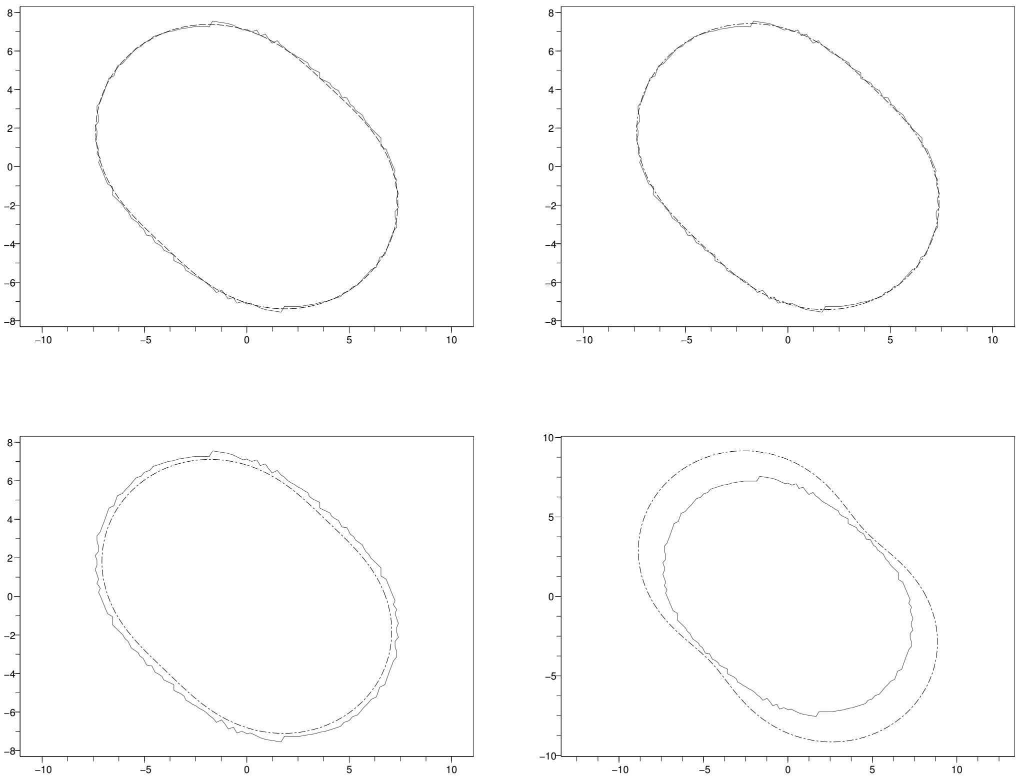


Figure 3: Full Gaussian distribution $\mathcal{N}(0, 1/2, 1/2, 1/6)$ for $\alpha = 0.995$. Top left: comparison between a numerical method and the use of the equivalent (3.1) for the computation of the iso-quantile curve, full line: numerical method, dashed line: asymptotic equivalent. Top right, bottom left and bottom right: best, median and worst estimates of the iso-quantile curve for $n = 200$, full line: numerical method, dashed-dotted line: estimator $\widehat{q}_n$.

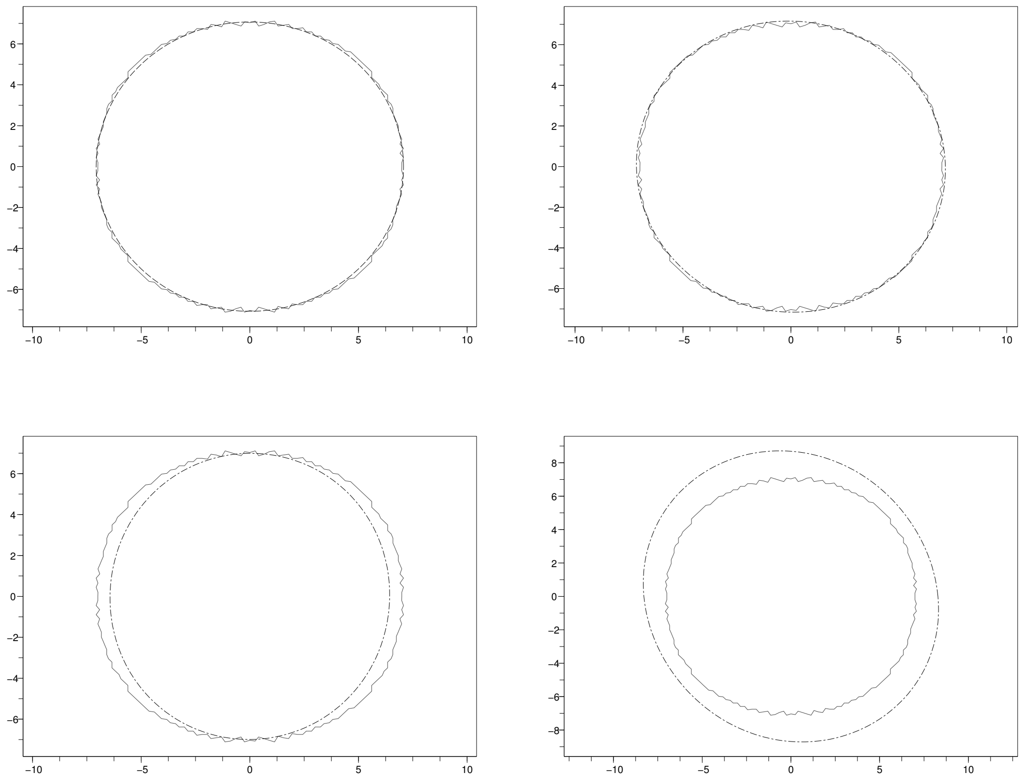


Figure 4: Spherical double exponential distribution $\mathcal{E}(2,2,2,2)$ for $\alpha = 0.995$. Top left: comparison between a numerical method and the use of the equivalent (3.1) for the computation of the iso-quantile curve, full line: numerical method, dashed line: asymptotic equivalent. Top right, bottom left and bottom right: best, median and worst estimates of the iso-quantile curve for $n = 200$, full line: numerical method, dashed-dotted line: estimator $\widehat{q}_n$.

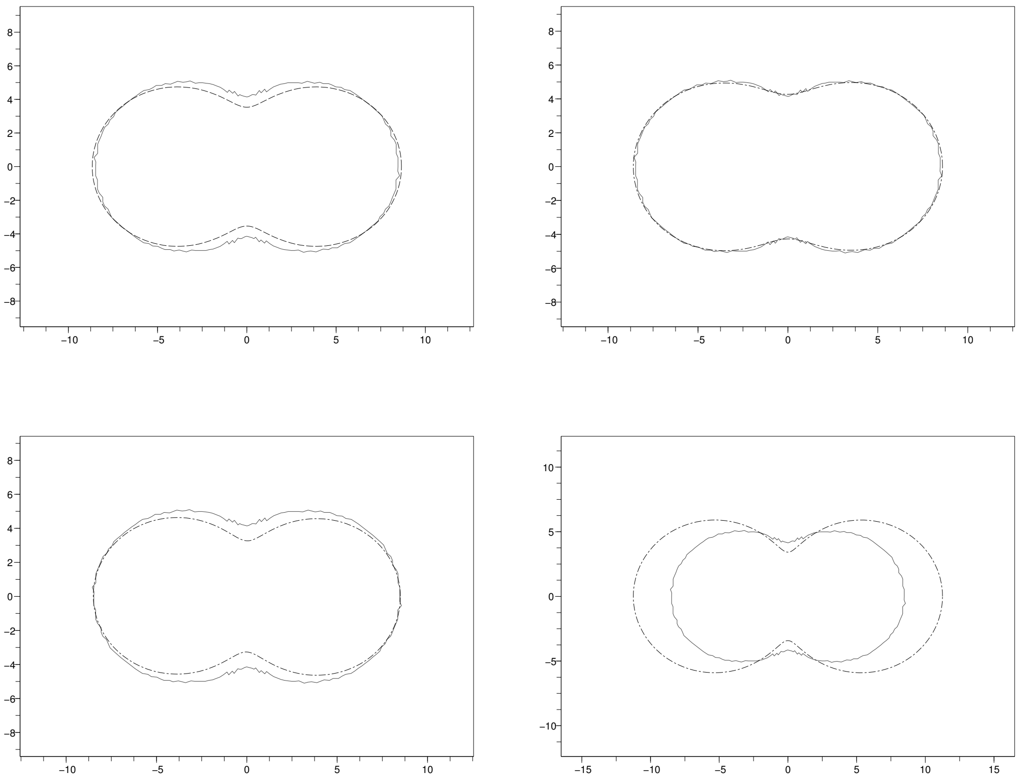


Figure 5: Diagonal double exponential distribution $\mathcal{E}(4, 2\sqrt{2/3}, 4, 2\sqrt{2/3})$ for $\alpha = 0.995$. Top left: comparison between a numerical method and the use of the equivalent (3.1) for the computation of the iso-quantile curve, full line: numerical method, dashed line: asymptotic equivalent. Top right, bottom left and bottom right: best, median and worst estimates of the iso-quantile curve for $n = 200$, full line: numerical method, dashed-dotted line: estimator $\widehat{q}_n$.

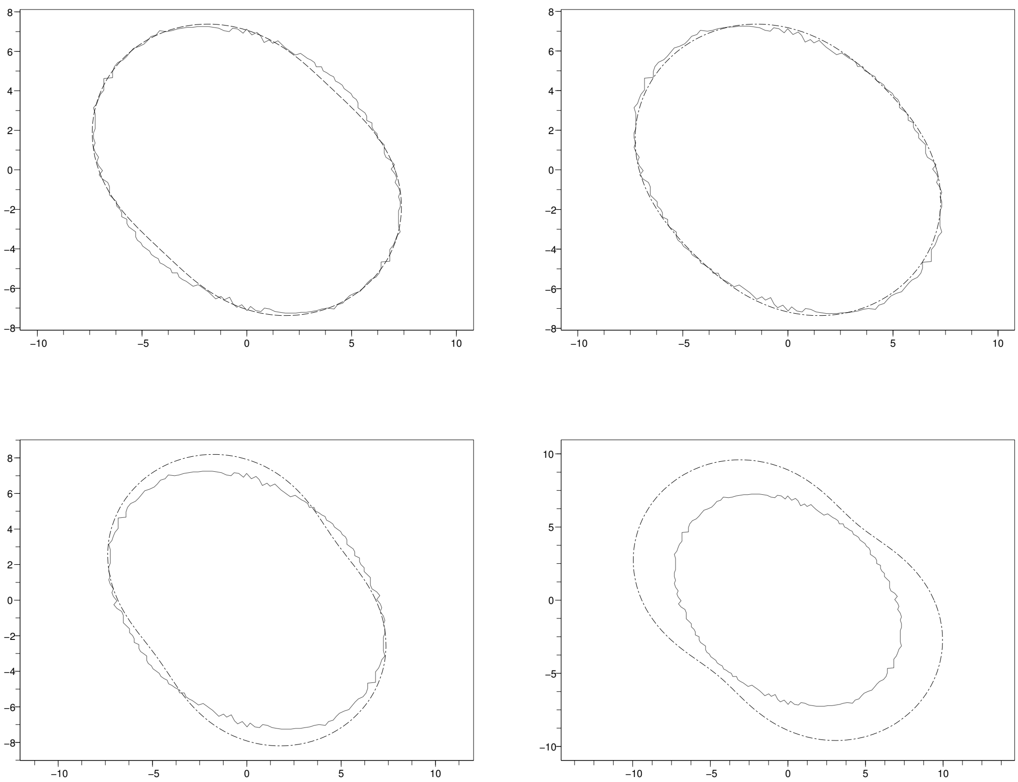


Figure 6: Full double exponential distribution $\mathcal{E}(2\sqrt{3}, 2\sqrt{3}, 2\sqrt{3/5}, 2\sqrt{3/5})$ for $\alpha = 0.995$. Top left: comparison between a numerical method and the use of the equivalent (3.1) for the computation of the iso-quantile curve, full line: numerical method, dashed line: asymptotic equivalent. Top right, bottom left and bottom right: best, median and worst estimates of the iso-quantile curve for $n = 200$, full line: numerical method, dashed-dotted line: estimator $\widehat{q}_n$.

Table 1: Errors $e(\alpha)$ and $E_n(\alpha)$ in all cases.

Distribution	Value of α	Error $e(\alpha)$	Error $E_n(\alpha)$		
			$n = 100$	$n = 200$	$n = 500$
Centred Gaussian $\mathcal{N}(0, 1/2, 1/2, 0)$	0.990	$2.55 \cdot 10^{-5}$	$1.29 \cdot 10^{-3}$	$6.50 \cdot 10^{-4}$	$2.93 \cdot 10^{-4}$
	0.995	$2.43 \cdot 10^{-5}$	$1.28 \cdot 10^{-3}$	$6.44 \cdot 10^{-4}$	$2.88 \cdot 10^{-4}$
	0.999	$5.75 \cdot 10^{-5}$	$1.30 \cdot 10^{-3}$	$6.70 \cdot 10^{-4}$	$3.16 \cdot 10^{-4}$
Centred Gaussian $\mathcal{N}(0, 1/2, 1/2, 1/6)$	0.990	$1.05 \cdot 10^{-4}$	$1.45 \cdot 10^{-3}$	$7.32 \cdot 10^{-4}$	$3.57 \cdot 10^{-4}$
	0.995	$4.34 \cdot 10^{-5}$	$1.37 \cdot 10^{-3}$	$6.65 \cdot 10^{-4}$	$2.89 \cdot 10^{-4}$
	0.999	$6.34 \cdot 10^{-5}$	$1.38 \cdot 10^{-3}$	$6.83 \cdot 10^{-4}$	$3.05 \cdot 10^{-4}$
Centred Gaussian $\mathcal{N}(0, 1/8, 3/4, 0)$	0.990	$6.05 \cdot 10^{-4}$	$1.79 \cdot 10^{-3}$	$1.17 \cdot 10^{-3}$	$8.23 \cdot 10^{-4}$
	0.995	$1.77 \cdot 10^{-4}$	$1.34 \cdot 10^{-3}$	$7.31 \cdot 10^{-4}$	$3.91 \cdot 10^{-4}$
	0.999	$5.96 \cdot 10^{-5}$	$1.20 \cdot 10^{-3}$	$6.02 \cdot 10^{-4}$	$2.70 \cdot 10^{-4}$
Double exponential $\mathcal{E}(2, 2, 2, 2)$	0.990	$9.30 \cdot 10^{-5}$	$2.69 \cdot 10^{-3}$	$1.47 \cdot 10^{-3}$	$6.37 \cdot 10^{-4}$
	0.995	$5.46 \cdot 10^{-5}$	$2.63 \cdot 10^{-3}$	$1.41 \cdot 10^{-3}$	$5.93 \cdot 10^{-4}$
	0.999	$6.32 \cdot 10^{-5}$	$2.63 \cdot 10^{-3}$	$1.39 \cdot 10^{-3}$	$5.97 \cdot 10^{-4}$
Double exponential $\mathcal{E}(2\sqrt{3}, 2\sqrt{3}, 2\sqrt{3/5}, 2\sqrt{3/5})$	0.990	$6.17 \cdot 10^{-4}$	$4.37 \cdot 10^{-3}$	$2.71 \cdot 10^{-3}$	$1.42 \cdot 10^{-3}$
	0.995	$2.24 \cdot 10^{-4}$	$3.89 \cdot 10^{-3}$	$2.26 \cdot 10^{-3}$	$9.96 \cdot 10^{-4}$
	0.999	$2.27 \cdot 10^{-4}$	$3.77 \cdot 10^{-3}$	$2.16 \cdot 10^{-3}$	$9.62 \cdot 10^{-4}$
Double exponential $\mathcal{E}(4, 2\sqrt{2/3}, 4, 2\sqrt{2/3})$	0.990	$1.64 \cdot 10^{-3}$	$4.13 \cdot 10^{-3}$	$2.81 \cdot 10^{-3}$	$2.16 \cdot 10^{-3}$
	0.995	$8.13 \cdot 10^{-4}$	$3.27 \cdot 10^{-3}$	$1.98 \cdot 10^{-3}$	$1.33 \cdot 10^{-3}$
	0.999	$6.62 \cdot 10^{-5}$	$2.40 \cdot 10^{-3}$	$1.23 \cdot 10^{-3}$	$5.62 \cdot 10^{-4}$

4.2. Real data illustration

The finite sample behaviour of extreme geometric quantiles is illustrated on a two-dimensional dataset extracted from the Pima Indians Diabetes Database. This data set[1] was already considered by [15] and [12], among others. In the latter study, geometric iso-quantile curves with a high α are used to detect outliers in the data set. Using extreme quantiles for outlier detection was advocated in *e.g.* [5, 20] in the univariate case and [19] using depth-based quantile regions in the multivariate case; see also the monograph [1].

After working on the data set so as to eliminate missing values, the data set consists of $n = 392$ pairs (X_i, Y_i), where X_i is the body mass index (BMI) of the i-th individual and Y_i is its diastolic blood pressure. The centered data cloud is represented in Figure 7 with blue crosses, along with the geometric iso-quantile curve with $\alpha = 0.95$. While geometric quantiles with a moderate α tend to give a fair idea of the shape of the data cloud (see *e.g.* [12]), the same cannot be said for extreme geometric quantiles on this example. This is an illustration of the phenomenon described in Consequence 3 in Section 2: the norm of an extreme geometric quantile is the largest in the direction where the variance is the smallest.

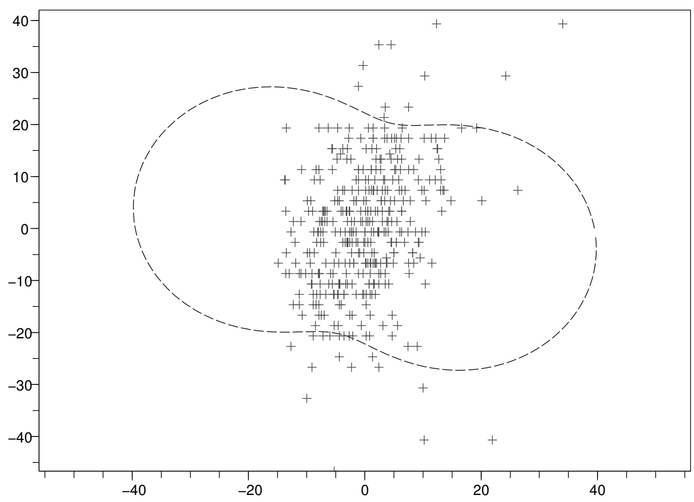


Figure 7: Pima Indians Diabetes data set. Black dashed line: estimate of the iso-quantile curve for $\alpha = 0.95$, with the estimator $\widehat{q}_n$.

We are thus led to think that here, outlier detection would be dangerous without a preliminary transformation–retransformation procedure [10].

[1]Available at `ftp.ics.uci.edu/pub/machine-learning-databases/pima-indians-diabetes`

5. CONCLUDING REMARKS

In this paper, we established the asymptotics of extreme geometric quantiles. A particular consequence of our results is that, if the underlying distribution possesses a finite covariance matrix Σ, then an extreme geometric quantile may be estimated accurately, no matter how extreme it is, with the help of the standard empirical estimator of Σ. This result is supported by our numerical study. The situation is very different from the univariate case, in which the asymptotic decay of a survival function can be linked to the asymptotic behaviour of an extreme quantile.

An additional issue, illustrated on a real data set, is that although central geometric quantile contours may roughly match the shape of the data cloud, this does not necessarily stay true for extreme iso-quantile curves. This is why we would advise practitioners to be cautious when using such a notion of multivariate quantile to detect outliers or analyze the extremes of a random vector. We believe that one can tackle this problem by applying a transformation–retransformation procedure, see [27] at the population level, and [9, 10] at the sample level. Future work on extreme geometric quantiles thus includes building and studying their analogues for transformed–retransformed data.

Finally, let us underline again that this work was carried out under moment conditions such as the existence of finite first and second-order moments for $\|X\|$. The case when these assumptions are violated is investigated in [17].

6. PROOFS

Some preliminary results are collected in Paragraph 6.1, their proofs are postponed to Paragraph 6.3. The proofs of the main results are provided in Paragraph 6.2.

6.1. Preliminary results

The first lemma provides some technical tools necessary to show Theorem 2.2(ii).

Lemma 6.1. *Let $\varphi\colon \mathbb{R}^d \times \mathbb{R}_+ \times S^{d-1} \to \mathbb{R}$ be the function defined by*

$$\varphi(x,r,v) \;=\; r^2 \left[1 + \frac{\langle x - rv,\, v\rangle}{\|x - rv\|}\right].$$

Then, for all $v \in S^{d-1}$, $\varphi(\cdot,\cdot,v)$ *is nonnegative and*

$$\forall x \in \mathbb{R}^d, \quad \forall r \le \|x\|, \quad \varphi(x,r,v) \le 2r^2 \quad \text{ and } \quad \forall r > \|x\|, \quad \varphi(x,r,v) \le \|x\|^2 .$$

In particular, $\varphi(x,r,v) \le 2\|x\|^2$ *for every* $(x,r,v) \in \mathbb{R}^d \times \mathbb{R}_+ \times S^{d-1}$.

The next lemma is the first step to prove Theorem 2.2(i).

Lemma 6.2. *Let* $u \in S^{d-1}$. *If* $\mathbb{E}\|X\| < \infty$ *then, for all* $v \in \mathbb{R}^d$,

$$\|q(\alpha u)\| \left\langle \alpha u - \frac{q(\alpha u)}{\|q(\alpha u)\|},\, v \right\rangle \to -\mathbb{E}\big\langle X - \langle X,u\rangle u,\, v\big\rangle \qquad \text{as} \quad \alpha \uparrow 1 .$$

Lemma 6.3 below is a result which is similar to Lemma 6.2.

Lemma 6.3. *Let* $u \in S^{d-1}$. *If* $\mathbb{E}\|X\|^2 < \infty$ *then*

$$\|q(\alpha u)\|^2 \left\langle \alpha u - \frac{q(\alpha u)}{\|q(\alpha u)\|},\, \frac{q(\alpha u)}{\|q(\alpha u)\|} \right\rangle \to -\frac{1}{2}\, \mathbb{E}\big\|X - \langle X,u\rangle u\big\|^2 \qquad \text{as} \quad \alpha \uparrow 1 .$$

Lemma 6.4 is the first step to prove Theorem 3.2. It is essentially a refinement of Lemma 6.2.

Lemma 6.4. *Let* $u \in S^{d-1}$. *If* $\mathbb{E}\|X\|^2 < \infty$ *then, for all* $v \in \mathbb{R}^d$,

$$\begin{aligned} &\|q(\alpha u)\| \left[\|q(\alpha u)\| \left\langle \alpha u - \frac{q(\alpha u)}{\|q(\alpha u)\|},\, v \right\rangle + \mathbb{E}\big\langle X - \langle X,u\rangle u,\, v\big\rangle \right] \to \\ &\quad \to \langle u,v\rangle \operatorname{Var}\langle X,u\rangle - \frac{1}{2}\langle u,v\rangle\, \mathbb{E}\big\|X - \langle X,u\rangle u\big\|^2 + \langle u,v\rangle \big\|\mathbb{E}\big(X - \langle X,u\rangle u\big)\big\|^2 \\ &\qquad - \operatorname{Cov}\big(\langle X,u\rangle, \langle X,v\rangle\big) \end{aligned}$$

as $\alpha \uparrow 1$.

Lemma 6.5 below is a refinement of Lemma 6.3. It is the second step to prove Theorem 3.2.

Lemma 6.5. *Let* $u \in S^{d-1}$. *If* $\mathbb{E}\|X\|^3 < \infty$ *then*

$$\begin{aligned} &\|q(\alpha u)\| \left(\|q(\alpha u)\|^2 \left\langle \alpha u - \frac{q(\alpha u)}{\|q(\alpha u)\|},\, \frac{q(\alpha u)}{\|q(\alpha u)\|} \right\rangle + \frac{1}{2}\, \mathbb{E}\big\|X - \langle X,u\rangle u\big\|^2 \right) \to \\ &\quad \to \mathbb{E}\Big(\langle X,u\rangle \Big[\big\langle X,\, \mathbb{E}\big(X - \langle X,u\rangle u\big) \big\rangle - \big\|X - \langle X,u\rangle u\big\|^2 \Big] \Big) \qquad \text{as} \quad \alpha \uparrow 1 . \end{aligned}$$

6.2. Proofs of the main results

Proof of Proposition 2.1: From [14], it is known that if $u \in B^d$ then problem (1.1) has a unique solution $q(u) \in \mathbb{R}^d$. To prove the converse part of this result, use equation (2.1) to get

$$\left\| \mathbb{E}\left(\frac{X - q(u)}{\|X - q(u)\|} \right) \right\| = \|u\| .$$

Let us introduce the coordinate representations $X = (X_1, ..., X_d)$ and $q(u) = (q_1(u), ..., q_d(u))$. The Cauchy–Schwarz inequality yields

$$\begin{aligned} \|u\|^2 &= \left\| \mathbb{E}\left(\frac{X - q(u)}{\|X - q(u)\|} \right) \right\|^2 = \sum_{i=1}^{d} \left[\mathbb{E}\left(\frac{X_i - q_i(u)}{\|X - q(u)\|} \right) \right]^2 \\ &\le \sum_{i=1}^{d} \mathbb{E}\left(\frac{(X_i - q_i(u))^2}{\|X - q(u)\|^2} \right) = 1 . \end{aligned}$$

Furthermore, equality holds if and only if for all $i \in \{1, ..., d\}$, there exists $\mu_i \in \mathbb{R}$ such that

$$\frac{X_i - q_i(u)}{\|X - q(u)\|} = \mu_i$$

almost surely. In particular, if $w = (\mu_1, ..., \mu_d)$, this entails $X \in D = q(u) + \mathbb{R}w$ almost surely, which cannot hold since the distribution of X is not concentrated in a single straight line in $\mathbb{R}^d$. It follows that necessarily $\|u\|^2 < 1$, which is the result. □

Proof of Proposition 2.2:

(i) Note that (2.1) implies that, for any linear isometry h of $\mathbb{R}^d$ and every $u \in B^d$,

$$h(u) + \mathbb{E}\left(\frac{h(X) - h \circ q(u)}{\|X - q(u)\|} \right) = 0 .$$

Since h is a linear isometry, the random vectors X and $h(X)$ have the same distribution and the equality $\|X - q(u)\| = \|h(X) - h \circ q(u)\|$ holds almost surely. It follows that

$$h(u) + \mathbb{E}\left(\frac{X - h \circ q(u)}{\|X - h \circ q(u)\|} \right) = 0 .$$

Since $h(u) \in B^d$, it follows that $h \circ q(u) = q \circ h(u)$, which completes the proof of the first statement.

(ii) To prove the second part of Proposition 2.2, start by noting that since X and $-X$ have the same distribution, it holds that $\mathbb{E}\,(X/\|X\|) = 0$. The case $u = 0$ is then obtained via (2.1). If $u \neq 0$, up to using the first part of the result with a suitable linear isometry, we shall assume without loss of generality that

$u = (u_1, 0, ..., 0)$ for some constant $u_1 \in (0,1)$. It is then enough to prove that there exists some constant $q_1(u) > 0$ such that $q(u) = (q_1(u), 0, ..., 0)$. To this end, let us remark that, on the one hand, if $v_1 \in \mathbb{R}$ and $w = (1, 0, ..., 0)$ then

$$\forall j \in \{2, ..., d\}, \quad \mathbb{E}\left(\frac{X_j}{\|X - v_1 w\|}\right) = 0 \,, \tag{6.1}$$

since, for all $j \in \{2, ..., d\}$, the random vectors $(X_1, ..., X_{j-1}, -X_j, X_{j+1}, ..., X_d)$ and X have the same distribution. On the other hand, the dominated convergence theorem entails that the function

$$v_1 \mapsto \mathbb{E}\left(\frac{X_1 - v_1}{\|X - v_1 w\|}\right)$$

is continuous, converges to 1 at $-\infty$, is equal to 0 at 0 and converges to -1 at $+\infty$. Thus, the intermediate value theorem yields that there exists some constant $q_1(u) > 0$ such that

$$u_1 + \mathbb{E}\left(\frac{X_1 - q_1(u)}{\|X - q_1(u) w\|}\right) = 0 \,. \tag{6.2}$$

Consequently, collecting (6.1) and (6.2) yields

$$u + \mathbb{E}\left(\frac{X - q_1(u) w}{\|X - q_1(u) w\|}\right) = 0$$

and it only remains to apply (2.1) to finish the proof of the second statement.

(iii) To show the third statement, use the first result to obtain that the function $g\colon \|u\| \mapsto \|q(u)\|$ is indeed well-defined; since the geometric quantile function is continuous, so is g. Assume that g is not strictly increasing: namely, there exist u_1, $u_2 \in B^d$ such that $\|u_1\| < \|u_2\|$ and $\|q(u_1)\| \geq \|q(u_2)\|$. Since $q(0) = 0$, it is a consequence of the intermediate value theorem that one may find u, $v \in B^d$ such that $\|u\| < \|v\|$ and $\|q(u)\| = \|q(v)\|$. Let h be an isometry such that $h(u/\|u\|) = h(v/\|v\|)$; then

$$\|q(h(u))\| = \|q(u)\| = \|q(v)\| = \|q(h(v))\|$$

and

$$\frac{q(h(u))}{\|q(h(u))\|} = \frac{h(u)}{\|h(u)\|} = \frac{h(v)}{\|h(v)\|} = \frac{q(h(v))}{\|q(h(v))\|} \,.$$

In other words, $q(h(u))$ and $q(h(v))$ have the same direction and magnitude, so that they are necessarily equal, which entails that $h(u) = h(v)$ because the geometric quantile function is one-to-one. This is a contradiction because $\|h(u)\| = \|u\| < \|v\| = \|h(v)\|$, and the third statement is proven.

(iv) Assume that $\|q(u)\|$ does not tend to infinity as $\|u\| \to 1$; since g is increasing, it tends to a finite positive limit r. In other words, $\|q(u)\| \leq r$ for every $u \in B^d$, which is a contradiction since the geometric quantile function maps B^d onto $\mathbb{R}^d$, and the proof is complete. $\square$

Proof of Theorem 2.1:

(**i**) If the first statement were false, then one could find a sequence (v_n) contained in B^d such that $\|v_n\| \to 1$ and such that $(\|q(v_n)\|)$ does not tend to infinity. Up to extracting a subsequence, one can assume that $(\|q(v_n)\|)$ is bounded. Again, up to extraction, one can assume that (v_n) converges to some $v_\infty \in S^{d-1}$ and that $(q(v_n))$ converges to some $q_\infty \in \mathbb{R}^d$. Moreover, it is straightforward to show that for every $u_1, u_2, q_1, q_2 \in \mathbb{R}^d$

$$\big|\psi(u_1, q_1) - \psi(u_2, q_2)\big| \;\le\; \big\{1 + \|u_2\|\big\}\, \|q_2 - q_1\| + \|q_1\|\, \|u_2 - u_1\|$$

so that the function ψ is continuous on $\mathbb{R}^d \times \mathbb{R}^d$. Recall then that the definition of $q(v_n)$ implies that for every $q \in \mathbb{R}^d$, $\psi(v_n, q(v_n)) \le \psi(v_n, q)$ and let n tend to infinity to obtain

$$q_\infty \;=\; \arg\min_{q \in \mathbb{R}^d} \psi(v_\infty, q)\;.$$

Because $v \in S^{d-1}$, this contradicts Proposition 2.1, and the proof of the first statement is complete: $\|q(v)\| \to \infty$ as $\|v\| \to 1$.

(**ii**) Pick a sequence (v_n) of elements of B^d converging to u and remark that from (2.1),

$$v_n + \mathbb{E}\left(\frac{X - q(v_n)}{\|X - q(v_n)\|}\right) \;=\; 0$$

for every integer n. Hence, for n large enough, the following equality holds:

$$v_n + \mathbb{E}\left(\left\|\frac{X}{\|q(v_n)\|} - \frac{q(v_n)}{\|q(v_n)\|}\right\|^{-1} \left[\frac{X}{\|q(v_n)\|} - \frac{q(v_n)}{\|q(v_n)\|}\right]\right) \;=\; 0\;. \tag{6.3}$$

Since the sequence $(q(v_n)/\|q(v_n)\|)$ is bounded it is enough to show that its only accumulation point is u. Let then u^* be an accumulation point of this sequence. Since $\|q(v_n)\| \to \infty$, we may let $n \to \infty$ in (6.3) and use the dominated convergence theorem to obtain $u - u^* = 0$, which completes the proof. □

Proof of Theorem 2.2:

(**i**) Let $(u, w_1, ..., w_{d-1})$ be an orthonormal basis of $\mathbb{R}^d$ and consider the following expansion:

$$\frac{q(\alpha u)}{\|q(\alpha u)\|} \;=\; b(\alpha)\, u + \sum_{k=1}^{d-1} \beta_k(\alpha)\, w_k \tag{6.4}$$

where $b(\alpha), \beta_1(\alpha), ..., \beta_{d-1}(\alpha)$ are real numbers. It immediately follows that

$$\frac{q(\alpha u)}{\|q(\alpha u)\|} - u - \frac{1}{\|q(\alpha u)\|}\big\{\mathbb{E}(X) - \langle \mathbb{E}(X), u\rangle\, u\big\} \;= \tag{6.5}$$

$$=\; \big(b(\alpha) - 1\big)\, u + \sum_{k=1}^{d-1} \frac{\|q(\alpha u)\|\beta_k(\alpha) - \mathbb{E}\langle X, w_k\rangle}{\|q(\alpha u)\|}\, w_k\;.$$

Lemma 6.2 implies that
(6.6)

$$\|q(\alpha u)\| \left\langle \alpha u - \frac{q(\alpha u)}{\|q(\alpha u)\|}, w_k \right\rangle = -\|q(\alpha u)\|\, \beta_k(\alpha) \;\to\; -\mathbb{E}\langle X, w_k\rangle \quad \text{as } \alpha \uparrow 1$$

for all $k \in \{1, ..., d-1\}$. Besides, let us note that $q(\alpha u)/\|q(\alpha u)\| \in S^{d-1}$ entails

$$b^2(\alpha) + \sum_{k=1}^{d-1} \beta_k^2(\alpha) \;=\; 1\,. \tag{6.7}$$

Theorem 2.1 shows that $b(\alpha) \to 1$ as $\alpha \uparrow 1$ and thus (6.6) yields:

$$\begin{aligned} \|q(\alpha u)\| \big(1 - b(\alpha)\big) &= \frac{1}{2}\, \|q(\alpha u)\| \big(1 - b^2(\alpha)\big) \big(1 + \mathrm{o}(1)\big) \\ &= \frac{1}{2}\, \|q(\alpha u)\| \sum_{k=1}^{d-1} \beta_k^2(\alpha) \big(1 + \mathrm{o}(1)\big) \;\to\; 0 \quad \text{as } \alpha \uparrow 1\,. \end{aligned} \tag{6.8}$$

Collecting (6.5), (6.6) and (6.8), we obtain

$$\frac{q(\alpha u)}{\|q(\alpha u)\|} - u - \frac{1}{\|q(\alpha u)\|} \big\{\mathbb{E}(X) - \langle \mathbb{E}(X), u\rangle\, u\big\} \;=\; \mathrm{o}\!\left(\frac{1}{\|q(\alpha u)\|}\right) \quad \text{as } \alpha \uparrow 1$$

which is the first result.

(**ii**) Recall (6.4) and use Lemma 6.2 to obtain

$$\|q(\alpha u)\| \left\langle \alpha u - \frac{q(\alpha u)}{\|q(\alpha u)\|}, w_k \right\rangle \to -\mathbb{E}\langle X, w_k\rangle \quad \text{as } \alpha \uparrow 1\,,$$

for all $k \in \{1, ..., d-1\}$, leading to

$$\|q(\alpha u)\|^2\, \beta_k^2(\alpha) \;\to\; \big[\mathbb{E}\langle X, w_k\rangle\big]^2 \quad \text{as } \alpha \uparrow 1 \tag{6.9}$$

for all $k \in \{1, ..., d-1\}$. Recall (6.7) and use Lemma 6.3 to get

$$\|q(\alpha u)\|^2 \big[\alpha b(\alpha) - 1\big] \;\to\; -\frac{1}{2}\, \mathbb{E}\big\|X - \langle X, u\rangle\, u\big\|^2 \quad \text{as } \alpha \uparrow 1\,. \tag{6.10}$$

Since $(u, w_1, ..., w_{d-1})$ is an orthonormal basis of $\mathbb{R}^d$, one has the identity

$$\big\|X - \langle X, u\rangle\, u\big\|^2 \;=\; \sum_{k=1}^{d-1} \langle X, w_k\rangle^2\,. \tag{6.11}$$

Collecting (6.9), (6.10) and (6.11) leads to

$$\|q(\alpha u)\|^2 \left[1 - \alpha b(\alpha) - \frac{1}{2} \sum_{k=1}^{d-1} \beta_k^2(\alpha)\right] \to \frac{1}{2} \sum_{k=1}^{d-1} \mathrm{Var}\langle X, w_k\rangle \quad \text{as } \alpha \uparrow 1\,.$$

Therefore,
(6.12)

$$\|q(\alpha u)\|^2 \left[1 - \alpha b(\alpha) - \frac{1}{2}\big(1 - b^2(\alpha)\big)\right] \to \frac{1}{2} \sum_{k=1}^{d-1} \mathrm{Var}\langle X, w_k\rangle \quad \text{as } \alpha \uparrow 1\,,$$

and easy calculations show that

$$1-\alpha\, b(\alpha)-\frac{1}{2}\left(1-b^2(\alpha)\right) \;=\; \frac{1}{2}\left[(1-\alpha)\,(1+\alpha)+\left(\alpha-b(\alpha)\right)^2\right]. \tag{6.13}$$

Finally, in view of Lemma 6.2,

$$\|q(\alpha u)\|\left\langle \alpha u-\frac{q(\alpha u)}{\|q(\alpha u)\|},\, u\right\rangle \;\to\; 0 \qquad \text{as}\;\; \alpha\uparrow 1$$

which is equivalent to

$$\|q(\alpha u)\|^2\left(\alpha-b(\alpha)\right)^2 \;\to\; 0 \qquad \text{as}\;\; \alpha\uparrow 1\,. \tag{6.14}$$

Collecting (6.12), (6.13) and (6.14), we obtain

$$\|q(\alpha u)\|^2\,(1-\alpha) \;\to\; \frac{1}{2}\sum_{k=1}^{d-1}\operatorname{Var}\langle X, w_k\rangle \qquad \text{as}\;\; \alpha\uparrow 1\,.$$

Remarking that, for every orthonormal basis $(e_1, ..., e_d)$ of $\mathbb{R}^d$,

$$\sum_{k=1}^{d}\operatorname{Var}\langle X, e_k\rangle \;=\; \sum_{k=1}^{d} e_k'\,\Sigma e_k \;=\; \operatorname{tr}\Sigma \tag{6.15}$$

proves that

$$\|q(\alpha u)\|^2\,(1-\alpha) \;\to\; \frac{1}{2}\left(\operatorname{tr}\Sigma-u'\Sigma u\right)\geq 0 \qquad \text{as}\;\; \alpha\uparrow 1\,.$$

Finally, note that if we had $\operatorname{tr}\Sigma-u'\Sigma u=0$ then by (6.15) we would have that $\operatorname{Var}\langle X, w_k\rangle=0$ for all $k\in\{1, ..., d-1\}$. Thus the projection of X onto the orthogonal complement of $\mathbb{R}u$ would be almost surely constant and X would be contained in a single straight line in $\mathbb{R}^d$, which is a contradiction. This completes the proof of Theorem 2.2. □

Proof of Theorem 3.1: Note that

$$\sqrt{1-\alpha_n}\;\widehat{q}_n(\alpha_n u) \;\to\; \left[\frac{1}{2}\left(\operatorname{tr}\Sigma-u'\Sigma u\right)\right]^{1/2} u \tag{6.16}$$

almost surely as $n\to\infty$. Moreover, by Theorems 2.1 and 2.2

$$\sqrt{1-\alpha_n}\;q(\alpha_n u) \;=\; \sqrt{1-\alpha_n}\;\|q(\alpha_n u)\|\,\frac{q(\alpha_n u)}{\|q(\alpha_n u)\|} \;\to\; \left[\frac{1}{2}\left(\operatorname{tr}\Sigma-u'\Sigma u\right)\right]^{1/2} u \tag{6.17}$$

almost surely as $n\to\infty$. Combining (6.16) and (6.17) completes the proof. □

Proof of Theorem 3.2: Consider the following representation:

$$\sqrt{n(1-\alpha_n)}\,\big(\widehat{q}_n(\alpha_n u) - q(\alpha_n u)\big) \;=\; T_{1,n} + T_{2,n} + T_{3,n}$$

with $$T_{1,n} \;=\; \sqrt{n}\left(\left[\frac{1}{2}\left\{\operatorname{tr}\widehat{\Sigma}_n - u'\widehat{\Sigma}_n u\right\}\right]^{1/2} - \left[\frac{1}{2}\left\{\operatorname{tr}\Sigma - u'\Sigma u\right\}\right]^{1/2}\right)\frac{q(\alpha_n u)}{\|q(\alpha_n u)\|}\,,$$

$$T_{2,n} \;=\; \sqrt{n}\left(\left[\frac{1}{2}\left\{\operatorname{tr}\Sigma - u'\Sigma u\right\}\right]^{1/2} - \sqrt{1-\alpha_n}\;\|q(\alpha_n u)\|\right)\frac{q(\alpha_n u)}{\|q(\alpha_n u)\|}$$

and $$T_{3,n} \;=\; -\sqrt{n(1-\alpha_n)}\;\|\widehat{q}_n(\alpha_n u)\|\left(\frac{q(\alpha_n u)}{\|q(\alpha_n u)\|} - u\right).$$

We start by examining the convergence of $T_{1,n}$. Observe first that

$$\begin{aligned} T_{1,n} \;&=\; \sqrt{n}\,\frac{1}{\sqrt{2}}\,\frac{\left\{\operatorname{tr}\widehat{\Sigma}_n - u'\widehat{\Sigma}_n u\right\} - \left\{\operatorname{tr}\Sigma - u'\Sigma u\right\}}{\left\{\operatorname{tr}\widehat{\Sigma}_n - u'\widehat{\Sigma}_n u\right\}^{1/2} + \left\{\operatorname{tr}\Sigma - u'\Sigma u\right\}^{1/2}}\,\frac{q(\alpha_n u)}{\|q(\alpha_n u)\|} \\ &=\; \sqrt{n}\,\frac{\left\{\operatorname{tr}\widehat{\Sigma}_n - u'\widehat{\Sigma}_n u\right\} - \left\{\operatorname{tr}\Sigma - u'\Sigma u\right\}}{2\sqrt{2}\,\left\{\operatorname{tr}\Sigma - u'\Sigma u\right\}^{1/2}}\,u\big(1+\mathrm{o}_{\mathbb{P}}(1)\big) \qquad \text{as } n\to\infty \end{aligned}$$

in view of Theorem 2.1(i) and from the consistency of $\widehat{\Sigma}_n$. Denote by M the Gaussian centred limit of $\sqrt{n}(\widehat{\Sigma}_n - \Sigma)$ (see *e.g.* [24]). Since the map $A \mapsto \operatorname{tr} A - u'Au$ is linear, it follows that

$$\sqrt{n}\,\frac{\left\{\operatorname{tr}\widehat{\Sigma}_n - u'\widehat{\Sigma}_n u\right\} - \left\{\operatorname{tr}\Sigma - u'\Sigma u\right\}}{2\sqrt{2}\,\left\{\operatorname{tr}\Sigma - u'\Sigma u\right\}^{1/2}} \;\xrightarrow{d}\; Y \qquad \text{as } n\to\infty$$

where Y is a centred Gaussian random variable. Now, clearly $Z := Yu$ is a Gaussian centred random vector and we have

$$T_{1,n} \xrightarrow{d} Z \qquad \text{as } n\to\infty\,. \tag{6.18}$$

The sequence $T_{2,n}$ is controlled in the following way: using Lemmas 6.4 and 6.5 and following the steps of the proof of Theorem 2.2(ii), we obtain

$$\begin{aligned} \|q(\alpha_n u)\|^2\,(1-\alpha_n) \;&=\; \frac{1}{2}\big(\operatorname{tr}\Sigma - u'\Sigma u\big) + \mathrm{O}\big(\|q(\alpha_n u)\|^{-1}\big) \\ &=\; \frac{1}{2}\big(\operatorname{tr}\Sigma - u'\Sigma u\big) + \mathrm{O}\big(\sqrt{1-\alpha_n}\,\big) \qquad \text{as } n\to\infty\,. \end{aligned}$$

As a consequence

$$\|T_{2,n}\| = \mathrm{O}\left(\sqrt{n(1-\alpha_n)}\right) = \mathrm{o}(1) \qquad \text{as } n\to\infty\,. \tag{6.19}$$

We conclude by controlling $T_{3,n}$. Theorem 2.2 entails

$$\begin{aligned} \|T_{3,n}\| \;&=\; \mathrm{O}_{\mathbb{P}}\left(\sqrt{n(1-\alpha_n)}\;\frac{\|\widehat{q}_n(\alpha_n u)\|}{\|q(\alpha_n u)\|}\right) \\ &=\; \mathrm{O}_{\mathbb{P}}\left(\sqrt{n(1-\alpha_n)}\;\left[\frac{\operatorname{tr}\widehat{\Sigma}_n - u'\widehat{\Sigma}_n u}{\operatorname{tr}\Sigma - u'\Sigma u}\right]^{1/2}\right) \\ &=\; \mathrm{O}_{\mathbb{P}}\left(\sqrt{n(1-\alpha_n)}\right) \;=\; \mathrm{o}_{\mathbb{P}}(1) \qquad \text{as } n\to\infty \end{aligned} \tag{6.20}$$

by the consistency of $\widehat{\Sigma}_n$. Combining (6.18), (6.19) and (6.20) completes the proof. □

6.3. Proofs of the preliminary results

Proof of Lemma 6.1: The fact that φ is nonnegative and the inequality

$$\forall r \le \|x\| \,, \qquad \varphi(x,r,v) \le 2\, r^2 \tag{6.21}$$

are straightforward consequences of the Cauchy–Schwarz inequality. Furthermore, φ can be rewritten as

$$\varphi(x,r,v) \;=\; r^2 \left[\frac{\big\|x - \langle x,v\rangle\, v\big\|^2}{\|x-rv\| \,\big[\|x-rv\| - \langle x-rv,\, v\rangle\big]} \right].$$

Let us now remark that, if $\|x\| < r$, then, by the Cauchy–Schwarz inequality, $\langle x - rv,\, v\rangle = \langle x, v\rangle - r < 0$ which makes it clear that

$$\varphi(x,r,v)\; 1\!\!1_{\{\|x\|<r\}} \;\le\; r^2\, \frac{\big\|x - \langle x,v\rangle\, v\big\|^2}{\|x-rv\|^2}\; 1\!\!1_{\{\|x\|<r\}} \;=:\; \psi(x,r,v)\; 1\!\!1_{\{\|x\|<r\}}\,. \tag{6.22}$$

Since $\|x-rv\|^2 = \|x\|^2 - 2r\langle x,\, v\rangle + r^2$, the function $\psi(x,\cdot,v)$ is differentiable on $(\|x\|,\, +\infty)$ and some easy computations yield

$$\frac{\partial \psi}{\partial r}(x,r,v) \;=\; 2\, r\, \big[\|x\|^2 - r\langle x,\, v\rangle\big]\, \frac{\big\|x - \langle x,v\rangle\, v\big\|^2}{\|x-rv\|^4}\,.$$

If $\langle x,\, v\rangle \le 0$ then $\psi(x,\cdot,v)$ is increasing on $(\|x\|,\, +\infty)$ and thus

$$\forall r > \|x\|\,, \qquad \psi(x,r,v) \le \lim_{r\to+\infty} \psi(x,r,v) \;=\; \|x - \langle x,v\rangle v\|^2 \;\le\; \|x\|^2\,. \tag{6.23}$$

Otherwise, if $\langle x, v\rangle > 0$ then $\psi(x,\cdot,v)$ reaches its global maximum over $[\|x\|, +\infty)$ at $\|x\|^2/\langle x,\, v\rangle$ and therefore,

$$\forall r > \|x\|\,, \qquad \psi(x,r,v) \;\le\; \psi\left(x,\, \frac{\|x\|^2}{\langle x, v\rangle},\, v\right) \;=\; \|x\|^2\,. \tag{6.24}$$

Collecting (6.22), (6.23) and (6.24) yields

$$\varphi(x,r,v)\; 1\!\!1_{\{\|x\|<r\}} \;\le\; \|x\|^2\; 1\!\!1_{\{\|x\|<r\}}\,. \tag{6.25}$$

Combining (6.21) and (6.25) shows that $\varphi(x,r,v) \le 2\|x\|^2$ for every $r > 0$ and every $v \in S^{d-1}$ and completes the proof of the result. □

Proof of Lemma 6.2: Let $v \in \mathbb{R}^d$ and $W_\alpha(\cdot, v)\colon \mathbb{R}^d \to \mathbb{R}$ be the function defined by

$$W_\alpha(x,v) \;=\; \left[\left\| \frac{x}{\|q(\alpha u)\|} - \frac{q(\alpha u)}{\|q(\alpha u)\|} \right\|^{-1} - 1 \right] \left\langle \frac{x}{\|q(\alpha u)\|} - \frac{q(\alpha u)}{\|q(\alpha u)\|},\, v \right\rangle.$$

For α close enough to 1, (2.1) entails

$$\left\langle \alpha u - \frac{q(\alpha u)}{\|q(\alpha u)\|},\, v \right\rangle + \mathbb{E}\big(W_\alpha(X,v)\big) + \frac{1}{\|q(\alpha u)\|}\,\mathbb{E}\langle X, v\rangle \;=\; 0\,. \tag{6.26}$$

It is therefore enough to show that

$$\|q(\alpha u)\|\,\mathbb{E}\big(W_\alpha(X,v)\big) \;\to\; -\langle u,\, v\rangle\,\mathbb{E}\langle X, u\rangle \qquad \text{as} \;\; \alpha \uparrow 1\,. \tag{6.27}$$

Since, for every $x \in \mathbb{R}^d$,

$$\left\| \frac{x}{\|q(\alpha u)\|} - \frac{q(\alpha u)}{\|q(\alpha u)\|} \right\|^2 \;=\; 1 - \frac{2}{\|q(\alpha u)\|}\left\langle x,\, \frac{q(\alpha u)}{\|q(\alpha u)\|}\right\rangle + \frac{\|x\|^2}{\|q(\alpha u)\|^2}\,, \tag{6.28}$$

it follows from a Taylor expansion and Theorem 2.1 that

$$\|q(\alpha u)\|\,W_\alpha(X,v) \;\to\; -\langle u, v\rangle\,\langle X, u\rangle \qquad \text{almost surely as} \;\; \alpha \uparrow 1\,. \tag{6.29}$$

Besides,

$$\begin{aligned} \left| \left\| \frac{x}{\|q(\alpha u)\|} - \frac{q(\alpha u)}{\|q(\alpha u)\|} \right\|^{-1} - 1 \right| \;=\; & \left\| \frac{x}{\|q(\alpha u)\|} - \frac{q(\alpha u)}{\|q(\alpha u)\|} \right\|^{-1} \\ & \times \left[1 + \left\| \frac{x}{\|q(\alpha u)\|} - \frac{q(\alpha u)}{\|q(\alpha u)\|} \right\| \right]^{-1} \\ & \times \left| \frac{2}{\|q(\alpha u)\|}\left\langle x,\, \frac{q(\alpha u)}{\|q(\alpha u)\|}\right\rangle - \frac{\|x\|^2}{\|q(\alpha u)\|^2} \right|\,, \end{aligned}$$

and the Cauchy–Schwarz inequality yields

$$\left\| \frac{x}{\|q(\alpha u)\|} - \frac{q(\alpha u)}{\|q(\alpha u)\|} \right\|^{-1} \left\langle \frac{x}{\|q(\alpha u)\|} - \frac{q(\alpha u)}{\|q(\alpha u)\|},\, v \right\rangle \;\le\; \|v\|\,.$$

Thus, using the triangular inequality and the Cauchy–Schwarz inequality, it follows that

$$|W_\alpha(x,v)| \;\le\; \|v\| \left[1 + \left\| \frac{x}{\|q(\alpha u)\|} - \frac{q(\alpha u)}{\|q(\alpha u)\|} \right\| \right]^{-1} \frac{\|x\|}{\|q(\alpha u)\|} \left[2 + \frac{\|x\|}{\|q(\alpha u)\|} \right].$$

Consequently, one has

$$\|q(\alpha u)\|\;|W_\alpha(x,v)|\;\mathbb{1}_{\{\|x\| \le \|q(\alpha u)\|\}} \;\le\; 3\,\|v\|\,\|x\|\,\mathbb{1}_{\{\|x\| \le \|q(\alpha u)\|\}}\,.$$

Furthermore, the reverse triangle inequality entails, for $x \in \mathbb{R}^d$ such that $\|x\| > \|q(\alpha u)\|$:

$$\left[1 + \left\| \frac{x}{\|q(\alpha u)\|} - \frac{q(\alpha u)}{\|q(\alpha u)\|} \right\| \right]^{-1} \le \frac{\|q(\alpha u)\|}{\|x\|}\,,$$

and therefore,

$$\|q(\alpha u)\|\;|W_\alpha(x,v)|\;\mathbb{1}_{\{\|x\| > \|q(\alpha u)\|\}} \;\le\; 3\,\|v\|\,\|x\|\,\mathbb{1}_{\{\|x\| > \|q(\alpha u)\|\}}\,.$$

Finally,

$$\|q(\alpha u)\|\;|W_\alpha(X,v)| \;\le\; 3\,\|v\|\,\|X\|$$

so that the integrand in (6.27) is bounded from above by an integrable random variable. One can now recall (6.29) and apply the dominated convergence theorem to obtain (6.27). The proof is complete. □

Proof of Lemma 6.3: Let $Z_\alpha\colon \mathbb{R}^d \to \mathbb{R}$ be the function defined by

$$Z_\alpha(x) \;=\; 1 + \left\langle \frac{x - q(\alpha u)}{\|x - q(\alpha u)\|},\, \frac{q(\alpha u)}{\|q(\alpha u)\|} \right\rangle .$$

For α close enough to 1, (2.1) yields

$$\left\langle \alpha u - \frac{q(\alpha u)}{\|q(\alpha u)\|},\, \frac{q(\alpha u)}{\|q(\alpha u)\|} \right\rangle + \mathbb{E}\big(Z_\alpha(X)\big) \;=\; 0 \tag{6.30}$$

and it thus remains to prove that

$$\|q(\alpha u)\|^2\, \mathbb{E}\big(Z_\alpha(X)\big) \;\to\; \frac{1}{2}\, \mathbb{E}\big\|X - \langle X, u\rangle\, u\big\|^2 \qquad \text{as} \;\; \alpha \uparrow 1 \,.$$

To this end, rewrite Z_α as

$$Z_\alpha(x) \;=\; 1 - \left\| \frac{x}{\|q(\alpha u)\|} - \frac{q(\alpha u)}{\|q(\alpha u)\|} \right\|^{-1} \left[1 - \frac{1}{\|q(\alpha u)\|} \left\langle x,\, \frac{q(\alpha u)}{\|q(\alpha u)\|} \right\rangle \right] . \tag{6.31}$$

It thus follows from equation (6.28), Theorem 2.1 and a Taylor expansion that

$$Z_\alpha(x) \;=\; \frac{1}{2\,\|q(\alpha u)\|^2} \left\langle x - \left\langle x,\, \frac{q(\alpha u)}{\|q(\alpha u)\|} \right\rangle \frac{q(\alpha u)}{\|q(\alpha u)\|},\, x \right\rangle \big(1 + \mathrm{o}(1)\big)$$

for all $x \in \mathbb{R}^d$. Using Theorem 2.1 again, we then get

$$\|q(\alpha u)\|^2\, Z_\alpha(X) \;\to\; \|X\|^2 - \langle X, u\rangle^2 \;=\; \|X - \langle X, u\rangle\, u\|^2 \qquad \text{almost surely as} \;\; \alpha \uparrow 1 \,. \tag{6.32}$$

To conclude the proof, let $\varphi\colon \mathbb{R}^d \times \mathbb{R}_+ \times S^{d-1} \to \mathbb{R}$ be the function defined by

$$\varphi(x, r, v) \;=\; r^2 \left[1 + \frac{\langle x - rv,\, v\rangle}{\|x - rv\|} \right] .$$

Note that $\|q(\alpha u)\|^2 Z_\alpha(x) = \varphi\big(x, \|q(\alpha u)\|, q(\alpha u)/\|q(\alpha u)\|\big)$. By Lemma 6.1:

$$\|q(\alpha u)\|^2\, Z_\alpha(X) \;=\; \varphi\Big(X,\, \|q(\alpha u)\|,\, q(\alpha u)/\|q(\alpha u)\|\Big) \;\le\; 2\,\|X\|^2$$

and the right-hand side is an integrable random variable. Use then (6.32) and the dominated convergence theorem to complete the proof. $\square$

Proof of Lemma 6.4: Let $v \in \mathbb{R}^d$ and recall the notation

$$W_\alpha(x, v) \;=\; \left[\left\| \frac{x}{\|q(\alpha u)\|} - \frac{q(\alpha u)}{\|q(\alpha u)\|} \right\|^{-1} - 1 \right] \left\langle \frac{x}{\|q(\alpha u)\|} - \frac{q(\alpha u)}{\|q(\alpha u)\|},\, v \right\rangle$$

from the proof of Lemma 6.2. From (6.26) there, it is enough to show that

$$\begin{aligned} \|q(\alpha u)\|\, \mathbb{E}\Big(\|q(\alpha u)\|\, W_\alpha(X, v) + \langle u, v\rangle\, \langle X, u\rangle \Big) \;\to& \\ \to\; \frac{1}{2}\, \langle u, v\rangle\, \mathbb{E}\big\|X - \langle X, u\rangle\, u\big\|^2 &- \langle u, v\rangle\, \mathrm{Var}\langle X, u\rangle \\ &+ \mathrm{Cov}\big(\langle X, u\rangle, \langle X, v\rangle\big) - \langle u, v\rangle\, \big\|\mathbb{E}(X - \langle X, u\rangle\, u)\big\|^2 \end{aligned} \tag{6.33}$$

as $\alpha \uparrow 1$. Use now (6.28) in the proof of Lemma 6.2, Theorem 2.2(i) and a Taylor expansion to obtain after some cumbersome computations that

$$\begin{aligned}
&\|q(\alpha u)\| \Big(\|q(\alpha u)\| \, W_\alpha(X,v) + \langle u,v\rangle \, \langle X,u\rangle \Big) \; = \\
&\quad = \frac{1}{2} \langle u,v\rangle \, \|X - \langle X,u\rangle u\|^2 - \langle u,v\rangle \, \langle X,u\rangle \, \big(\langle X,u\rangle - \mathbb{E}\langle X,u\rangle\big) \\
&\qquad + \langle X,u\rangle \, \big(\langle X,v\rangle - \mathbb{E}\langle X,v\rangle\big) - \langle u,v\rangle \, \big\langle X, \, \mathbb{E}\big(X - \langle X,u\rangle u\big)\big\rangle \\
&\qquad + \sum_{j=0}^{2} \|X\|^j \, \varepsilon_j\big(\alpha, X, q(\alpha u)\big)
\end{aligned}$$

with probability 1, where for all $j \in \{0,1,2\}$, $\varepsilon_j(\alpha,y,z) \to 0$ as $\max(1-\alpha, \|y\|/\|z\|) \downarrow 0$. In particular

$$\begin{aligned}
&\|q(\alpha u)\| \Big(\|q(\alpha u)\| \, W_\alpha(X,v) + \langle u,v\rangle \, \langle X,u\rangle \Big) \; \to \\
&\quad \to \frac{1}{2} \langle u,v\rangle \, \big\|X - \langle X,u\rangle u\big\|^2 - \langle u,v\rangle \, \langle X,u\rangle \, \big(\langle X,u\rangle - \mathbb{E}\langle X,u\rangle\big) \\
&\qquad - \langle u,v\rangle \, \big\langle X, \, \mathbb{E}\big(X - \langle X,u\rangle u\big)\big\rangle + \langle X,u\rangle \, \big(\langle X,v\rangle - \mathbb{E}\langle X,v\rangle\big) \, ,
\end{aligned} \tag{6.34}$$

almost surely as $\alpha \uparrow 1$. The proof shall be complete provided we can apply the dominated convergence theorem to the left-hand side of (6.34). To this end, let $\delta \in (0,1)$ be such that

$$\alpha \in (1-\delta, 1) \;\text{ and }\; \frac{\|X\|}{\|q(\alpha u)\|} < \delta \implies \max_{0 \le j \le 2} \big|\varepsilon_j(\alpha, X, q(\alpha u))\big| \le 1 \, .$$

Equality (6.34) thus entails for α close enough to 1:

$$\begin{aligned}
&\|q(\alpha u)\| \, \Big| \|q(\alpha u)\| \, W_\alpha(X,v) + \langle u,v\rangle \, \langle X,u\rangle \Big| \, \mathbb{1}_{\{\|X\| < \delta \|q(\alpha u)\|\}} \; \le \\
&\qquad\qquad \le \; P_1\big(\|X\|\big) \, \mathbb{1}_{\{\|X\| < \delta \|q(\alpha u)\|\}}
\end{aligned}$$

where P_1 is a real polynomial of degree 2. Besides, it is a consequence of the definition of $W_\alpha(X,v)$ and the Cauchy–Schwarz inequality that

$$\begin{aligned}
&\|q(\alpha u)\| \, \Big| \|q(\alpha u)\| \, W_\alpha(X,v) + \langle u,v\rangle \, \langle X,u\rangle \Big| \, \mathbb{1}_{\{\|X\| \ge \delta \|q(\alpha u)\|\}} \; \le \\
&\qquad\qquad \le \; \frac{2\,(1+\delta)\,\|v\|}{\delta^2} \, \|X\|^2 \, \mathbb{1}_{\{\|X\| \ge \delta \|q(\alpha u)\|\}} \, .
\end{aligned}$$

One can conclude that there exists a real polynomial P_2 of degree 2 such that

$$\|q(\alpha u)\| \, \Big| \|q(\alpha u)\| \, W_\alpha(X,v) + \langle u,v\rangle \, \langle X,u\rangle \Big| \; \le \; P_2\big(\|X\|\big)$$

so that the integrand in (6.33) is bounded by an integrable random variable. Recall (6.34) and apply the dominated convergence theorem to complete the proof. $\square$

Proof of Lemma 6.5: The proof is similar to that of Lemma 6.4. Recall from the proof of Lemma 6.3 the notation

$$Z_\alpha(x) \;=\; 1 + \left\langle \frac{x - q(\alpha u)}{\|x - q(\alpha u)\|},\, \frac{q(\alpha u)}{\|q(\alpha u)\|} \right\rangle .$$

From (6.30) there, it is enough to show that

$$\text{(6.35)} \qquad \|q(\alpha u)\|\; \mathbb{E}\Big(\|q(\alpha u)\|^2\, Z_\alpha(X) - \frac{1}{2}\, \mathbb{E}\big\|X - \langle X,u\rangle\, u\big\|^2 \Big) \;\to$$
$$\to\; \mathbb{E}\Big(\langle X,u\rangle\, \Big[\big\|X - \langle X,u\rangle\, u\big\|^2 - \big\langle X,\, \mathbb{E}\big(X - \langle X,u\rangle u\big)\big\rangle \Big] \Big)$$

as $\alpha \uparrow 1$. We first use (6.28) in the proof of Lemma 6.2, equation (6.31) in the proof of Lemma 6.3, Theorem 2.2(i) and a Taylor expansion to obtain after some burdensome computations that

$$\text{(6.36)} \qquad \|q(\alpha u)\|\; \Big(\|q(\alpha u)\|^2\, Z_\alpha(X) - \frac{1}{2}\, \big\|X - \langle X,u\rangle\, u\big\|^2 \Big) \;=$$
$$=\; \langle X,u\rangle\, \Big(\big\|X - \langle X,u\rangle\, u\big\|^2 - \big\langle X,\, \mathbb{E}\big(X - \langle X,u\rangle\, u\big)\big\rangle \Big) + \sum_{j=0}^{3} \|X\|^j\, \varepsilon_j\big(\alpha, X, q(\alpha u)\big)$$

with probability 1, where for $j \in \{0,1,2,3\}$, $\varepsilon_j(\alpha, y, z) \to 0$ as $\max(1-\alpha,$ $\|y\|/\|z\|) \downarrow 0$. Especially

$$\text{(6.37)} \qquad \|q(\alpha u)\|\; \Big(\|q(\alpha u)\|^2\, Z_\alpha(X) - \frac{1}{2}\, \big\|X - \langle X,u\rangle\, u\big\|^2 \Big) \;\to$$
$$\to\; \langle X,u\rangle\, \Big(\big\|X - \langle X,u\rangle\, u\big\|^2 - \big\langle X,\, \mathbb{E}\big(X - \langle X,u\rangle\, u\big)\big\rangle \Big)$$

as $\alpha \uparrow 1$. Our aim is now to apply the dominated convergence theorem to the left-hand side of (6.35). To this end, pick $\delta \in (0,1)$ such that

$$\alpha \in (1-\delta, 1) \;\text{ and }\; \frac{\|X\|}{\|q(\alpha u)\|} < \delta \;\Longrightarrow\; \max_{0\le j\le 3} \big|\varepsilon_j\big(\alpha, X, q(\alpha u)\big)\big| \;\le\; 1 \,.$$

Equality (6.36) thus entails for α close enough to 1:

$$\|q(\alpha u)\|\, \Big| \|q(\alpha u)\|^2\, Z_\alpha(X) - \frac{1}{2}\, \big\|X - \langle X,u\rangle\, u\big\|^2 \Big|\, \mathbb{1}_{\{\|X\| < \delta \|q(\alpha u)\|\}} \;\le$$
$$\le\; P_1\big(\|X\|\big)\, \mathbb{1}_{\{\|X\| < \delta \|q(\alpha u)\|\}}$$

where P_1 is a real polynomial of degree 3. Moreover, the Cauchy–Schwarz inequality yields

$$\|q(\alpha u)\|\, \Big| \|q(\alpha u)\|^2\, Z_\alpha(X) - \frac{1}{2}\, \|X - \langle X,u\rangle\, u\|^2 \Big|\, \mathbb{1}_{\{\|X\| \ge \delta \|q(\alpha u)\|\}} \;\le$$
$$\le\; \frac{4+\delta^2}{2\,\delta^3}\; \|X\|^3\, \mathbb{1}_{\{\|X\| \ge \delta \|q(\alpha u)\|\}} \,.$$

Consequently, there exists a real polynomial P_2 of degree 3 such that

$$\|q(\alpha u)\| \left| \|q(\alpha u)\|^2 Z_\alpha(X) - \frac{1}{2} \left\| X - \langle X, u\rangle u \right\|^2 \right| \le P_2\big(\|X\|\big).$$

We conclude that the integrand in (6.35) is bounded by an integrable random variable. Recall (6.37) and apply the dominated convergence theorem to complete the proof. □

REFERENCES

[1] Aggarwal, C. C. (2013). *Outlier analysis*, Springer, New York.

[2] Bingham, N. H.; Goldie, C. M. and Teugels, J. L. (1987). *Regular Variation*, Cambridge, U.K.: Cambridge University Press.

[3] Breckling, J. and Chambers, R. (1988). M-quantiles, *Biometrika*, **75**(4), 761–771.

[4] Breckling, J.; Kokic, P. and Lübke, O. (2001). A note on multivariate M-quantiles, *Stat. Probab. Lett.*, **55**, 39–44.

[5] Burridge, P. and Robert Taylor, A. M. (2006). Additive outlier detection via extreme-value theory, *J. Time Series Anal.*, **27**(5), 685–701.

[6] Cadre, B. and Gannoun, A. (2000). Asymptotic normality of consistent estimate of the conditional L_1-median, *Ann. I.S.U.P.*, **44**, 13–35.

[7] Cambanis, S.; Huang, S. and Simons, G. (1981). On the theory of elliptically contoured distributions, *J. Multivariate Anal.*, **11**(3), 368–385.

[8] Cardot, H.; Cénac, P. and Zitt, P.-A. (2013). Efficient and fast estimation of the geometric median in Hilbert spaces with an averaged stochastic gradient algorithm, *Bernoulli*, **19**(1), 18–43.

[9] Chakraborty, B.; Chaudhuri, P. and Oja, H. (1998). Operating transformation retransformation on spatial median and angle test, *Statist. Sinica*, **8**, 767–784.

[10] Chakraborty, B. (2001). On affine equivariant multivariate quantiles, *Ann. Inst. Statist. Math.*, **53**(2), 380–403.

[11] Chakraborty, B. (2003). On multivariate quantile regression, *J. Statist. Plann. Inference*, **110**(2), 109–132.

[12] Chaouch, M. and Goga, C. (2010). Design-based estimation for geometric quantiles with application to outlier detection, *Comput. Statist. Data Anal.*, **54**, 2214–2229.

[13] Chaouch, M. and Laïb, N. (2013). Nonparametric multivariate L_1-median regression estimation with functional covariates, *Electron. J. Stat.*, **7**, 1553–1586.

[14] Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data, *J. Amer. Statist. Assoc.*, **91**(434), 862–872.

[15] Cheng, Y. and De Gooijer, J. G. (2007). On the u-th geometric conditional quantile, *J. Statist. Plann. Inference*, **137**, 1914–1930.

[16] Dhar, S. S.; Chakraborty, B. and Chaudhuri, P. (2014). Comparison of multivariate distributions using quantile-quantile plots and related tests, *Bernoulli*, **20**(3), 1484–1506.

[17] Girard, S. and Stupfler, G. (2015). Extreme geometric quantiles in a multivariate regular variation framework, *Extremes*, **18**, 629–663.

[18] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*, Springer, New York.

[19] He, Y. and Einmahl, J. H. J. (2016). Estimation of extreme depth-based quantile regions, *Journal of the Royal Statistical Society: Series B*, to appear.

[20] Huberta, M.; Dierckx, G. and Vanpaemela, D. (2012). Detecting influential data points for the Hill estimator in Pareto-type distributions, *Comput. Statist. Data Anal.*, **65**, 13–28.

[21] Kemperman, J. H. B. (1987). *The median of a finite measure on a Banach space*. In "Statistical Data Analysis Based on the L^1-norm and Related Methods" (Y. Dodge, Ed.), 217–230, North Holland, Amsterdam.

[22] Koenker, R. and Bassett, G. (1978). Regression quantiles, *Econometrica*, **46**, 33–50.

[23] Koltchinskii, V. I. (1997). M-estimation, convexity and quantiles, *Ann. Statist.*, **25**(2), 435–477.

[24] Neudecker, H. and Wesselman, A. M. (1990). The asymptotic variance matrix of the sample correlation matrix, *Linear Algebra Appl.*, **127**, 589–599.

[25] Serfling, R. (2002). Quantile functions for multivariate analysis: approaches and applications, *Stat. Neerl.*, **56**(2), 214–232.

[26] Serfling, R. (2004). Nonparametric multivariate descriptive measures based on spatial quantiles, *J. Statist. Plann. Inference*, **56**(2), 214–232.

[27] Serfling, R. (2010). Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardization, *J. Nonparametr. Stat.*, **22**, 915–936.

[28] Small, C. G. (1990). A survey of multidimensional medians, *Int. Stat. Rev.*, **58**(3), 263–277.

[29] Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations, *J. Amer. Statist. Assoc.*, **73**, 812–815.

REVSTAT – Statistical Journal
Volume 15, Number 1, January 2017, 141–153

THE SHORTEST CLOPPER–PEARSON RANDOMIZED CONFIDENCE INTERVAL FOR BINOMIAL PROBABILITY

Author: Wojciech Zieliński
– Department of Econometrics and Statistics,
Warsaw University of Life Sciences,
Nowoursynowska 159, 02-776 Warszawa, Poland
wojciech_zielinski@sggw.pl

Received: March 2014 Revised: June 2015 Accepted: July 2015

Abstract:

- Zieliński (2010) showed the existence of the shortest Clopper–Pearson confidence interval for binomial probability. The method of obtaining such an interval was presented as well. Unfortunately, the confidence interval obtained has one disadvantage: it does not keep the prescribed confidence level. In this paper, a small modification is introduced, after which the resulting shortest confidence interval does not have the above mentioned disadvantage.

Key-Words:

- *binomial proportion; confidence interval; shortest confidence intervals.*

AMS Subject Classification:

- 62F25.

1. INTRODUCTION

The problem of estimating the probability of success in a binomial model has a very long history as well as very wide applications. Let us recall the definition of a confidence interval for probability of success $\theta \in (0,1)$ (see Cramér (1946), Lehmann (1959), Silvey (1970); for a general definition of confidence interval see Neyman (1934)):

A random interval $\big(\underline{\theta}(X), \overline{\theta}(X)\big)$ is called a *confidence interval for a parameter θ at the confidence level γ* if

$$P_\theta\Big\{\underline{\theta}(X) \le \theta \le \overline{\theta}(X)\Big\} \;\ge\; \gamma \qquad \text{for all} \quad \theta \in (0,1)\ .$$

Here X denotes the number of successes in a sample of size n.

It is easy to note that for any $d(n), g(n) > 0$ the interval $\big(\underline{\theta}(X) - d(n), \overline{\theta}(X) + g(n)\big)$ is of course also a confidence interval. So an additional criterion is needed for choosing a confidence interval. There are a lot different criterions. Clopper and Pearson, who proposed the first confidence interval for θ, took under consideration the equal risk of underestimation and overestimation. The problem of the shortest confidence intervals was seldom considered in the past (Crow 1956, Blyth and Hutchinson 1960, Blyth and Still 1983, Casella 1986). Zieliński (2010) proposed a simple method of obtaining the shortest confidence interval for θ. Unfortunately, the solution has a serious disadvantage: the proposed confidence interval does not keep the nominal confidence level. So, in what follows a slight modification is proposed. Namely, an auxiliary random variable $Y \in (0,1)$ is applied and the shortest confidence interval is constructed on the basis of $X+Y$. It appears that such a confidence interval does not have the above mentioned disadvantage.

2. THE CONFIDENCE INTERVAL

Consider the binomial statistical model

$$\Big(\big\{0,1,...,n\big\},\ \big\{Bin(n,\theta),\ 0<\theta<1\big\}\Big)\ ,$$

where $Bin(n,\theta)$ denotes the binomial distribution with probability distribution function (pdf)

$$\binom{n}{k}\theta^k(1-\theta)^{n-k}\ , \qquad k = 0,1,...,n\ .$$

It is well known that

$$\sum_{k\le x}\binom{n}{k}\theta^k(1-\theta)^{n-k} \;=\; F\big(n-x,\, x+1;\, 1-\theta\big) \;=\; 1 - F\big(x+1,\, n-x;\, \theta\big)\ ,$$

where $F(a,b;\cdot)$ is the cumulative distribution function (cdf) of the beta distribution with parameters (a,b).

Let X denote a binomial $Bin(n,\theta)$ random variable. A confidence interval for probability θ at the confidence level γ is of the form (Clopper and Pearson, 1934)

$$\Big(F^{-1}(X, n-X+1; \gamma_1);\ F^{-1}(X+1, n-X; \gamma_2)\Big),$$

where $\gamma_1, \gamma_2 \in (0,1)$ are such that $\gamma_2 - \gamma_1 = \gamma$ and $F^{-1}(a,b;\alpha)$ is the α quantile of the beta distribution with parameters (a,b), i.e.

$$P_\theta\Big\{\theta \in \Big(F^{-1}\big(X, n-X+1; \gamma_1\big);\ F^{-1}\big(X+1, n-X; \gamma_2\big)\Big)\Big\} \geq \gamma, \quad \forall\,\theta \in (0,1).$$

For $X = 0$ the left end is taken to be 0, and for $X = n$ the right end is taken to be 1.

Zieliński (2010) considered the length of the confidence interval when $X = x$ is observed:

$$d(\gamma_1, x) = F^{-1}\big(x+1, n-x; \gamma+\gamma_1\big) - F^{-1}\big(x, n-x+1; \gamma_1\big).$$

Let x be given. The existence as well as the method of finding $0 < \gamma_1^* < 1-\gamma$ such that $d(\gamma_1^*, x)$ is minimal was shown. Examples of shortest confidence intervals (*left*, *right*) are given in Table 1.

Table 1: The shortest c.i. ($n = 20$, $\gamma = 0.95$).

x	γ_1^*	*left*	*right*
0	0.00000	0.00000	0.13911
1	0.00000	0.00000	0.21611
2	0.00125	0.00261	0.28393
3	0.00561	0.01839	0.34998
4	0.00966	0.04318	0.41249
5	0.01302	0.07344	0.47156
6	0.01587	0.10763	0.52766
7	0.01840	0.14496	0.58118
8	0.02071	0.18496	0.63234
9	0.02288	0.22733	0.68126
10	0.02500	0.27196	0.72804

By symmetry, for $x > n/2$ we have $\gamma_1^*(x) = (1-\gamma) - \gamma_1^*(n-x)$, $\mathit{left}(x) = 1 - \mathit{right}(n-x)$ and $\mathit{right}(x) = 1 - \mathit{left}(n-x)$. The confidence level of the shortest confidence interval for probability θ equals

$$\sum_{x=0}^{n} \binom{n}{x} \theta^x (1-\theta)^{n-x}\, \mathbf{1}(x,\theta),$$

where

$$\mathbf{1}(x,\theta) \;=\; \begin{cases} 1 & \text{if } \theta \in \big(\mathit{left}(x), \mathit{right}(x)\big)\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

For $n = 20$ and $\gamma = 0.95$ the confidence level is shown in Figure 1.

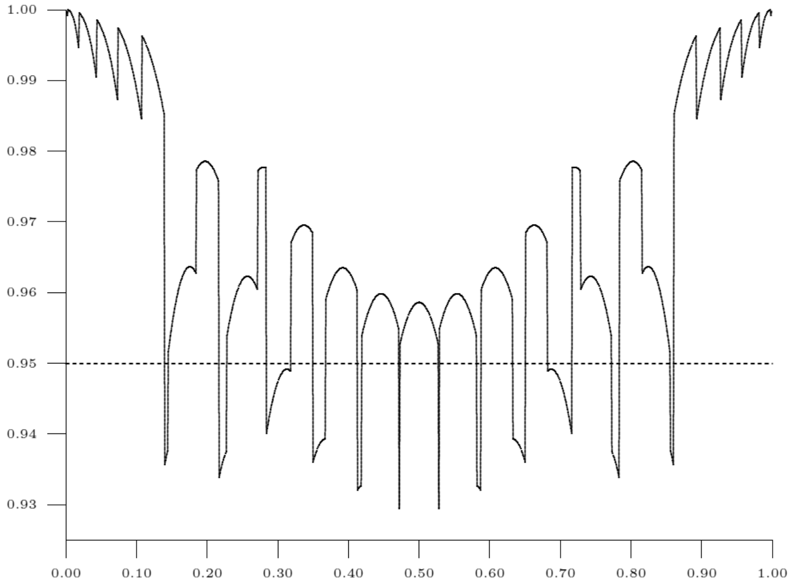


Figure 1: The confidence level of the shortest confidence interval: $n = 20$, $\gamma = 0.95$.

Note that for some probabilities θ the confidence level is smaller than the nominal one. This is in contradiction with the definition of the confidence interval (see Neyman 1934, Cramér 1949, Lehmann 1959, Silvey 1970).

In what follows, a small modification is introduced, after which the resulting shortest confidence interval does not have the above mentioned disadvantage, i.e. its confidence level is not smaller than the prescribed one.

Let Y be a random variable conditionally distributed on the interval $(0,1)$ with cdf $G_{Y|X=x}(\cdot)$. The confidence interval will be constructed on the basis of two random variables: $Z_s = X + Y$ and $Z_d = X - (1 - Y)$. The distributions of those r.v.'s are easy to obtain:

$$P_\theta\{Z_s \le t\} \;=\; \begin{cases} 0 & \text{if } t \le 0\,, \\ \alpha\big(\lfloor t\rfloor, \lceil t\rceil\big)\, P_\theta\{X = \lfloor t\rfloor\} & \text{if } \lfloor t\rfloor = 0\,, \\ \sum_{k=0}^{\lfloor t\rfloor - 1} P_\theta\{X = k\} + \alpha\big(\lfloor t\rfloor, \lceil t\rceil\big)\, P_\theta\{X = \lfloor t\rfloor\} & \text{if } 1 \le \lfloor t\rfloor \le n\,, \\ 1 & \text{if } \lfloor t\rfloor > n\,, \end{cases}$$

$$P_\theta\big\{Z_d \le t\big\} = \begin{cases} 0 & \text{if } t \le -1\,, \\ \alpha\big(\lfloor t\rfloor, \lceil t\rceil\big)\, P_\theta\big\{X = \lfloor t\rfloor + 1\big\} & \text{if } \lfloor t\rfloor = -1\,, \\ \sum_{k=0}^{\lfloor t\rfloor} P_\theta\big\{X = k\big\} + \alpha\big(\lfloor t\rfloor + 1, \lceil t\rceil\big)\, P_\theta\big\{X = \lfloor t\rfloor + 1\big\} & \text{if } 0 \le \lfloor t\rfloor \le n-1\,, \\ 1 & \text{if } \lfloor t\rfloor \ge n\,, \end{cases}$$

where $\lfloor t\rfloor$ denotes the greatest integer no greater than t and

$$\lceil t\rceil = t - \lfloor t\rfloor \qquad \text{and} \qquad \alpha\big(\lfloor t\rfloor, \lceil t\rceil\big) = \int_0^{\lceil t\rceil} G_{Y|X=\lfloor t\rfloor}(du)\,.$$

It is easy to note that the distribution of Y may be taken as the uniform $U(0,1)$ independently of X.

The shortest confidence interval (θ_L, θ_U) at the confidence level γ will be obtained as a solution with respect to θ of the following problem:

$$\begin{cases} \theta_U - \theta_L = \min!\,, \\ P_{\theta_L}\big\{Z_s \le t\big\} = \gamma_2\,, \\ P_{\theta_U}\big\{Z_d \ge t\big\} = 1 - \gamma_1\,, \\ \gamma_2 - \gamma_1 = \gamma\,. \end{cases}$$

Hence, for observed $X = x$ and $Y = y$ we have to find θ_L and θ_U such that

$$\begin{cases} \theta_U - \theta_L = \min!\,, \\ \sum_{k=0}^{x-1} P_{\theta_L}\big\{X = k\big\} + y P_{\theta_L}\big\{X = x\big\} = \gamma_2\,, \\ \sum_{k=0}^{x} P_{\theta_U}\big\{X = k\big\} + y\, P_{\theta_U}\big\{X = x+1\big\} = \gamma_1\,, \\ \gamma_2 - \gamma_1 = \gamma\,, \end{cases}$$

or, equivalently,

$$\begin{cases} \theta_U - \theta_L = \min!\,, \\ (1-y)\, F\big(x, n-x+1; \theta_L\big) + y F\big(x+1, n-x; \theta_L\big) = \gamma_1\,, \\ (1-y)\, F\big(x+1, n-x; \theta_U\big) + y F\big(x+2, n-x-1; \theta_U\big) = \gamma + \gamma_1\,. \end{cases}$$

Let

$$G(\theta; n, x, y) = (1-y)\, F\big(x, n-x+1; \theta\big) + y F\big(x+1, n-x; \theta\big)\,.$$

We take $F(a, 0; \theta) = 0$ and $F(0, b; \theta) = 1$. Then

$$\theta_L = G^{-1}(\gamma_1; n, x, y) \qquad \text{and} \qquad \theta_U = G^{-1}(\gamma + \gamma_1; n, x+1, y)\,.$$

In what follows we consider only the case $x \le n/2$. If $x \ge n/2$, the role of success and failure should be interchanged.

The problem of finding the shortest confidence interval may be written as the problem of finding γ_1 which minimizes

$$d(\gamma_1; n, x, y) \;=\; G^{-1}(\gamma+\gamma_1; n, x+1, y) - G^{-1}(\gamma_1; n, x, y)$$

for given $y \in [0,1]$, n and x.

Theorem 2.1. *For $x \geq 2$ and for all $y \in (0,1)$ there exists a two-sided shortest confidence interval.*

Proof: We have to show that for $x \geq 2$ and for all $y \in (0,1)$ there exists $0 < \gamma_1 < 1-\gamma$ such that $d(\gamma_1; n, x, y)$ is minimal. The derivative of $d(\gamma_1; n, x, y)$ with respect to γ_1 equals

$$\frac{\partial d(\gamma_1; n, x, y)}{\partial \gamma_1} \;=\; \frac{1}{LHS(\gamma_1; n, x, y)} - \frac{1}{RHS(\gamma_1; n, x, y)}$$

where

$$\begin{aligned} LHS(\gamma_1; n, x, y) \;=\; & \Big(1 - G^{-1}\big(\gamma+\gamma_1; n, x+1, y\big)\Big)^{n-x-1} G^{-1}\big(\gamma+\gamma_1; n, x+1, y\big)^{x+1} \\ & \cdot \Bigg(\frac{1-y}{G^{-1}(\gamma+\gamma_1; n, x+1, y)\, B(x+1, n-x)} \\ & \quad + \; \frac{y}{\big(1 - G^{-1}(\gamma+\gamma_1; n, x+1, y)\big)\, B(x+2, n-x-1)} \Bigg) \end{aligned}$$

and

$$\begin{aligned} RHS(\gamma_1; n, x, y) \;=\; & \Big(1 - G^{-1}(\gamma_1; n, x, y)\Big)^{n-x} G^{-1}(\gamma_1; n, x, y)^{x} \\ & \cdot \Bigg(\frac{1-y}{G^{-1}(\gamma_1; n, x, y)\, B(x, n-x+1)} \\ & \quad + \; \frac{y}{\big(1 - G^{-1}(\gamma_1; n, x, y)\big)\, B(x+1, n-x)} \Bigg). \end{aligned}$$

Because

$$G^{-1}(0; n, x, y) = 0 \qquad \text{and} \qquad G^{-1}(1; n, x, y) = 1\,,$$

for $2 \leq x \leq n/2$ we have:

$$\begin{array}{llll} \text{if } \; \gamma_1 \to 0 & \text{then } \; LHS(\gamma_1; n, x, y) > 0 & \text{and } \; RHS(\gamma_1; n, x, y) \to 0^+\,, \\ \text{if } \; \gamma_1 \to 1-\gamma & \text{then } \; LHS(\gamma_1; n, x, y) \to 0^+ & \text{and } \; RHS(\gamma_1; n, x, y) > 0\,. \end{array}$$

Therefore, the equation

$$(*) \qquad \frac{\partial d(\gamma_1; n, x, y)}{\partial \gamma_1} \;=\; 0$$

has a solution. □

It is easy to see that $LHS(\cdot;n,x,y)$ and $RHS(\cdot;n,x,y)$ are unimodal and concave on the interval $(0,1-\gamma)$. Hence, the solution of $(*)$ is unique. Let γ_1^* denote the solution. Because $\frac{\partial d(\gamma_1;n,x,y)}{\partial\gamma_1}<0$ for $\gamma_1<\gamma_1^*$ and $\frac{\partial d(\gamma_1;n,x,y)}{\partial\gamma_1}>0$ for $\gamma_1>\gamma_1^*$, we have $d(\gamma_1^*;n,x,y)=\inf\big\{d(\gamma_1;n,x,y)\colon 0<\gamma_1<1-\gamma\big\}$.

Theorem 2.2. *For $x=1$ there exists $y^*\in(0,1)$ such that if $Y<y^*$ then the shortest confidence interval is one-sided, and it is two-sided otherwise.*

Proof: For $x=1$ we have

$$\frac{\partial d(\gamma_1;n,1,y)}{\partial\gamma_1}=\frac{1}{LHS(\gamma_1;n,1,y)}-\frac{1}{RHS(\gamma_1;n,1,y)}$$

where

$$\begin{aligned}
LHS(\gamma_1;n,1,y) &= \Big(1-G^{-1}\big(\gamma+\gamma_1;n,2,y\big)\Big)^{n-2}G^{-1}\big(\gamma+\gamma_1;n,2,y\big)^2\\
&\quad\cdot\Bigg(\frac{1-y}{G^{-1}(\gamma+\gamma_1;n,2,y)\,B(2,n-2)}\\
&\qquad+\ \frac{y}{\big(1-G^{-1}(\gamma+\gamma_1;n,2,y)\big)\,B(3,n-2)}\Bigg)\\
&= \frac{1}{2}\,(n-1)\,n\big(1-G^{-1}(\gamma+\gamma_1;n,2,y)\big)^{n-3}G^{-1}(\gamma+\gamma_1;n,2,y)\\
&\quad\cdot\Big(2\big(1-G^{-1}(\gamma+\gamma_1;n,2,y)\big)+y\big(n\,G^{-1}(\gamma+\gamma_1;n,2,y)-2\big)\Big)
\end{aligned}$$

and

$$\begin{aligned}
RHS(\gamma_1;n,1,y) &= \Big(1-G^{-1}(\gamma_1;n,1,y)\Big)^{n-1}G^{-1}(\gamma_1;n,1,y)\\
&\quad\cdot\Bigg(\frac{1-y}{G^{-1}(\gamma_1;n,1,y)\,B(1,n)}\\
&\qquad+\ \frac{y}{\big(1-G^{-1}(\gamma_1;n,1,y)\big)\,B(2,n-1)}\Bigg)\\
&= n\big(1-G^{-1}(\gamma_1;n,1,y)\big)^{n-2}\\
&\quad\cdot\Big(1-G^{-1}(\gamma_1;n,1,y)+y\big(n\,G^{-1}(\gamma_1;n,1,y)-1\big)\Big)\,.
\end{aligned}$$

It can be seen that if $\gamma_1\to 0$, then

$$\begin{aligned}
LHS(\gamma_1;n,1,y) &\to \frac{\big(1-G^{-1}(\gamma;n,2,y)\big)^{n-3}G^{-1}(\gamma;n,2,y)}{B(2,n-1)}\\
&\quad\cdot\Big(2\big(1-G^{-1}(\gamma;n,2,y)\big)+y\big(n\,G^{-1}(\gamma;n,2,y)-2\big)\Big)\,,\\
RHS(\gamma_1;n,1,y) &\to (1-y)\,n\,.
\end{aligned}$$

If $\gamma_1\to 1-\gamma$, then

$$LHS(\gamma_1;n,1,y)\to 0^+\,,$$
$$RHS(\gamma_1;n,1,y)>0\,.$$

Because $LHS(0; n, 1, 0) < RHS(0; n, 1, 0)$ and $LHS(0; n, 1, 1) > RHS(0; n, 1, 1)$, there exists y^* such that $LHS(0; n, 1, y^*) = RHS(0; n, 1, y^*)$. So, for $y < y^*$ the shortest confidence interval is one-sided, and it is two-sided otherwise. $\square$

The value of y^* may be found numerically as a solution of

$$LHS(0; n, 1, y^*) = RHS(0; n, 1, y^*) .$$

In Table 2 the values of y^* for different n and confidence levels γ are given.

Table 2: Values of y^*.

n	$\gamma = 0.9$	$\gamma = 0.95$	$\gamma = 0.99$	$\gamma = 0.999$
10	0.783163	0.870995	0.964326	0.994792
20	0.828155	0.904080	0.976758	0.997138
30	0.840388	0.912599	0.979647	0.997620
40	0.846076	0.916491	0.980924	0.997825
50	0.849360	0.918718	0.981643	0.997937
60	0.851499	0.920160	0.982104	0.998009
70	0.853002	0.921170	0.982424	0.998058
80	0.854117	0.921917	0.982660	0.998094
90	0.854976	0.922491	0.982840	0.998121
100	0.855658	0.922947	0.982983	0.998143
150	0.857680	0.924294	0.983403	0.998206
200	0.858678	0.924955	0.983608	0.998237
300	0.859666	0.925610	0.983810	0.998267
400	0.860156	0.925934	0.983910	0.998281
500	0.860449	0.926128	0.983969	0.998290
600	0.860644	0.926257	0.984009	0.998296
700	0.860784	0.926349	0.984037	0.998300
800	0.860888	0.926418	0.984058	0.998303
900	0.860969	0.926471	0.984075	0.998306
1000	0.861034	0.926514	0.984088	0.998308

The above considerations may be summarized as follows. Observe a r.v. X distributed as $Bin(n, \theta)$ and draw Y distributed as $U(0, 1)$. If $X > n/2$ then consider $X' = n - X$. Calculate y^*, the solution of the equation $LHS(y^*; 0, n, 1) = RHS(y^*; 0, n, 1)$.

If $X + Y \leq 1 + y^*$ then the confidence interval is of the form

$$\Big(0;\, G^{-1}(\gamma; n, X + 1, Y)\Big) .$$

If $X + Y \geq 1 + y^*$ then find γ_1^* which minimizes $d(\gamma_1; n, x, y)$. Then the confidence interval takes on the form

$$\Big(G^{-1}(\gamma_1^*; n, X, Y);\, G^{-1}(\gamma + \gamma_1^*; n, X + 1, Y)\Big) .$$

If $X > n/2$ is observed then the shortest confidence interval has the form

$$\begin{cases} \Big(1-G^{-1}(\gamma;n,X'+1,Y);1\Big) & \text{if } X'+Y\le 1+y^*\,, \\ \Big(1-G^{-1}(\gamma+\gamma_1^*;n,X'+1,Y);\ 1-G^{-1}(\gamma_1^*;n,X',Y)\Big) & \text{otherwise}\,. \end{cases}$$

Theorem 2.3. *$P_\theta\{\theta_L \le \theta \le \theta_U\} \ge \gamma$ for $\theta \in (0,1)$, and $P_{0.5}\{\theta_L \le 0.5 \le \theta_U\} = \gamma$.*

Proof: Let $\theta \in (0,1)$ be given. Let x_u, y_u and γ_u be such that $\theta = G^{-1}(\gamma+\gamma_u; n, x_u, y_u)$. Similarly, let x_d, y_d and γ_d be such that $\theta = G^{-1}(\gamma_d; n, x_d+1, y_d)$. Of course, $x_d < x_u$ and $\gamma_d \le \gamma_u$. So

$$P_\theta\{\theta_L \le \theta \le \theta_U\} \;=\; P_\theta\{x_d + y_d \le X+Y \le x_u + y_u\} \;=\; \gamma + \gamma_u - \gamma_d \;\ge\; \gamma\,.$$

If $\theta = 0.5$ then, by symmetry, $x_u = n - x_d$, $y_u = 1 - y_d$ and $\gamma_u = \gamma_d$. Hence $P_{0.5}\{\theta_L \le 0.5 \le \theta_U\} = \gamma$. □

The confidence level of the randomized shortest confidence interval for $n = 20$ and $\gamma = 0.95$ is shown in Figure 2.

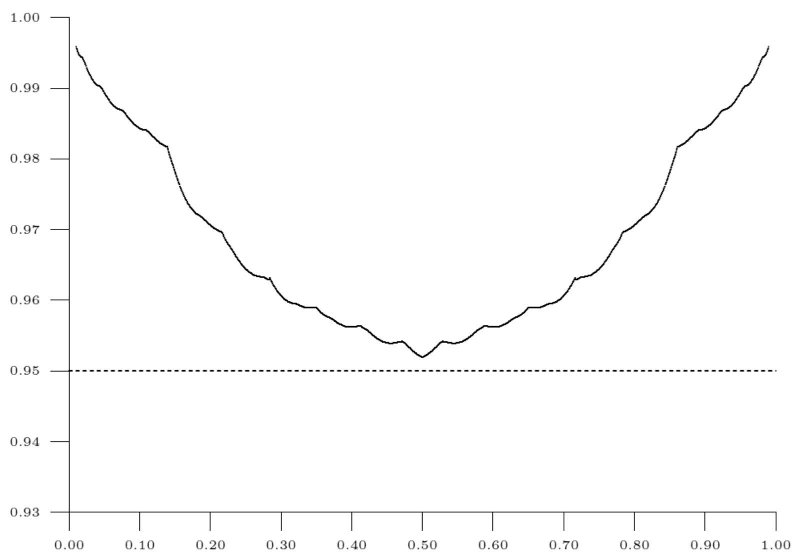


Figure 2: The confidence level of the randomized shortest confidence interval: $n = 20$, $\gamma = 0.95$.

In Clopper and Pearson's times, calculating quantiles of a beta distribution was numerically complicated. Nowadays, it is very easy with the aid of computer software, so using the shortest confidence interval is recommended (a short Mathematica program is given in the Appendix). To avoid problems with wrong inference due to the confidence level, one should use randomized shortest confidence intervals. Of course, the generated value y of a $U(0,1)$ r.v. must be attached to the final report. So results now are given by three numbers: number of trials, number of successes and the value y.

3. AN EXAMPLE

Consider an experiment consisting of $n = 20$ Bernoulli trials in which $x = 3$ successes were observed. Let $\gamma = 0.95$. The standard Clopper–Pearson confidence interval $\big(F^{-1}(3,18;0.025);\ F^{-1}(4,17;0.975)\big)$ takes on the form

$$(0.0321,\ 0.3789)\ .$$

The length of the standard Clopper–Pearson confidence interval equals 0.3468.

To calculate the randomized shortest confidence interval one has to draw a value y of the auxiliary variable Y and then calculate the ends of the confidence interval on the basis of $x + y$. The uniform random number generator gives $y = 0.0102162$ and the randomized shortest confidence interval takes on the form

$$(0.0184,\ 0.2898)\ .$$

The length of that confidence interval is 0.2714. Note that the length of the proposed confidence interval equals 78% of the length of the standard confidence interval.

The final report may look as follows:

$$n = 20\,,\quad x = 3\,,\quad y = 0.0102162\,,\quad \gamma = 0.95\,,\quad \theta \in (0.0184, 0.2898)\ .$$

In practical applications it is important to have conclusions as precise as possible. Hence the use of the randomized shortest confidence intervals is recommended, especially for small sample sizes. Those intervals are very easy to obtain with the aid of the standard computer software (see Appendix).

APPENDIX

Below we give a short Mathematica program for calculating γ_1^* and the ends of the randomized shortest confidence interval. Of course, one can also use other mathematical or statistical packages (in a similar way) to find the values of γ_1^*.

```
In[1]:= << Statistics'ContinuousDistributions'
      n=.; x=.; y=.; q=.;
      Bet[a_,b_,x_]=CDF[BetaDistribution[a, b], x];
      G[θ_,n_,x_,y_]=(1-y)*If[x==0,0,Bet[x,n-x+1,θ]]+y*If[x==n,0,Bet[x+1,n-x,θ]];
(*definition of the confidence interval*)
      Lower[p_,n_,x_,y_]:=If[x<=1+Ystar,0,θ/.FindRoot[G[θ,n,x,y]==p,{θ,0.001,0.999}];
      Upper[p_,n_,x_,y_]:=If[x>=n-1-Ystar,1,θ/.FindRoot[G[θ,n,x,y]==p,{θ,0.001,0.999}];
      Length[p_,n_,x_,y_,γ_]:=Upper[γ+p,n,x+1,y]-Lower[p,n,x,y];
In[2]:= n=20;(*input n*)
      x=7;(*input x (≤n/2)*)
      q=0.95 ;(*input confidence level*)
      y=RandomReal[];
(*calculate Y star*)
      eps=10^(-10); al=0; ar=1;
      While[ar-al>eps,{
       aa=(ar+al)/2;
       Dol=θ/.FindRoot[G[θ,n,2,aa]==q, {θ, 0.001, 0.999}];
       LHS=(1-Dol)^(n-3)*Dol*(2*(1-Dol)+aa*(n*Dol-2))/Beta[2,n-1];
       RHS=(1-aa)*n;
       If[LHS>RHS,ar=aa,al=aa];}]
      Ystar=aa;
(*calculate ends of the shortest confidence interval*)
      pp=If[x+y<=1+Ystar,0,
        p/.FindMinimum[Length[p,n,x,y,q],{p,0,1-q }][[2]]] (*probability γ1*)
      Left=Lower[pp,n,x,y] (*left end*)
      Right=Upper[q+pp,n,x,y] (*right end*)
      y (*drawn U(0,1) r.v.*)
```

REFERENCES

[1] Blyth, C.R. and Hutchinson, D.W. (1960). Table of Neyman-shortest unbiased confidence intervals for the binomial parameter, *Biometrika*, **47**, 381–391.

[2] Blyth, C.R. and Still, H.A. (1983). Binomial confidence intervals, *Journal of the American Statistical Association*, **78**, 108–116.

[3] Clopper, C.J. and Pearson, E.S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial, *Biometrika*, **26**, 404–413.

[4] Crow, J.L. (1956). Confidence intervals for a proportion, *Biometrika*, **45**, 423–435.

[5] Neyman, J. (1934). On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection, *Journal of the Royal Statistical Society*, **97**, 558–625.

[6] Cramér, H. (1946). *Mathematical Methods in Statistics*, Princeton University Press (Nineteenth printing 1999).

[7] Lehmann, E.L. (1959). *Testing Statistical Hypothesis*, Springer (Third edition 2005).

[8] Silvey, S.D. (1970). *Statistical Inference*, Chapman & Hall (Eleventh edition 2003).

[9] Zieliński, W. (2010). The shortest Clopper–Pearson confidence interval for binomial probability, *Communications in Statistics — Simulation and Computation*, **39**, 188–193, DOI: 10.1080/03610910903391270.

REVSTAT – STATISTICAL JOURNAL

Background

Statistics Portugal (INE, I.P.), well aware of how vital a statistical culture is in understanding most phenomena in the present-day world, and of its responsibility in disseminating statistical knowledge, started the publication of the scientific statistical journal *Revista de Estatística*, in Portuguese, publishing three times a year papers containing original research results, and application studies, namely in the economic, social and demographic fields.

In 1998 it was decided to publish papers also in English. This step has been taken to achieve a larger diffusion, and to encourage foreign contributors to submit their work.

At the time, the Editorial Board was mainly composed by Portuguese university professors, being now composed by national and international university professors, and this has been the first step aimed at changing the character of *Revista de Estatística* from a national to an international scientific journal.

In 2001, the *Revista de Estatística* published three volumes special issue containing extended abstracts of the invited contributed papers presented at the 23rd European Meeting of Statisticians.

The name of the Journal has been changed to REVSTAT - STATISTICAL JOURNAL, published in English, with a prestigious international editorial board, hoping to become one more place where scientists may feel proud of publishing their research results.

— The editorial policy will focus on publishing research articles at the highest level in the domains of Probability and Statistics with emphasis on the originality and importance of the research.

— All research articles will be refereed by at least two persons, one from the Editorial Board and another external.

— The only working language allowed will be English. — Four volumes are scheduled for publication, one in January, one in April, one in July and the other in October.

Aims and Scope

The aim of REVSTAT is to publish articles of high scientific content, in English, developing innovative statistical scientific methods and introducing original research, grounded in substantive problems.

REVSTAT covers all branches of Probability and Statistics. Surveys of important areas of research in the field are also welcome.

Abstract and Indexing Services

The REVSTAT is covered by the following abstracting/indexing services:

- Current Index to Statistics
- Google Scholar
- Mathematical Reviews
- Science Citation Index Expanded
- Zentralblatt für Mathematic

Instructions to Authors, special-issue editors and publishers

The articles should be written in English and may be submitted in two different ways:

- By sending the paper in PDF format to the Executive Editor (revstat@ine.pt) and to one of the two Editors or Associate Editors, whose opinion the author wants to be taken into account, together to the following e-mail address: revstat@fc.ul.pt

- By sending the paper in PDF format to the Executive Editor (revstat@ine.pt), together with the corresponding PDF or PostScript file to the following e-mail address: revstat@fc.ul.pt.

Submission of a paper means that it contains original work that has not been nor is about to be published elsewhere in any form.

Manuscripts (text, tables and figures) should be typed only in black on one side, in double-spacing, with a left margin of at least 3 cm and with less than 30 pages. The first page should include the name, institution and address of the author(s) and a summary of less than one hundred words, followed by a maximum of six key words and the AMS 2000 subject classification.

Authors are obliged to write the final version of accepted papers using LaTeX, in the REVSTAT style. This style (REVSTAT.sty), and examples file (REVSTAT.tex), which may be download to PC Windows System (Zip format), Macintosh, Linux and Solaris Systems (StuffIt format), and Mackintosh System (BinHex Format), are available in the REVSTAT link of the Statistics Portugal Website: http://www.ine.pt/revstat/inicio.html

Additional information for the authors may be obtained in the above link.

Accepted papers

Authors of accepted papers are requested to provide the LaTeX files and also a postscript (PS) or an acrobat (PDF) file of the paper to the Secretary of REVSTAT: revstat@ine.pt.

Such e-mail message should include the author(s)'s name, mentioning that it has been accepted by REVSTAT.

The authors should also mention if encapsulated postscript figure files were included, and submit electronics figures separately in .tiff, .gif, .eps or .ps format. Figures must be a minimum of 300 dpi.